



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2017

Dissecting the genetic architecture of GnRH deficiency using whole-exome sequencing

CASSATELLA Daniele

CASSATELLA Daniele, 2017, Dissecting the genetic architecture of GnRH deficiency using whole-exome sequencing

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_392A901D89943

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de Physiologie

**Dissecting the genetic architecture of GnRH deficiency
using whole-exome sequencing**

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Daniele CASSATELLA

Master en Biotechnologie Médicale de l'Université de Bari, Italie

Jury

Prof. Manuela Eicher, Président
Prof. Nelly Pitteloud, Directeur de thèse
Dr. Brian J. Stevenson, Co-directeur
Prof. Jacques Beckmann, expert
Prof. Alexandre Reymond, expert

Lausanne 2017



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président· e	Madame Prof. Manuela Eicher
Directeur· rice de thèse	Madame Prof. Nelly Pitteloud
Co-directeur· rice	Monsieur Dr Brian J. Stevenson
Experts· es	Monsieur Prof. Jacques Beckmann
	Monsieur Prof. Alexandre Reymond

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Daniele Cassatella

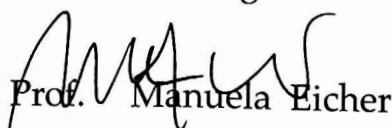
Master in Medical Biotechnologies and Molecular Medicine de l' Université de Bari, Italie

intitulée

**Dissecting the genetic architecture of GnRH deficiency
using whole-exome sequencing**

Lausanne, le 3 mai 2017

pour le Doyen
de la Faculté de biologie et de médecine



Prof. Manuela Eicher

Acknowledgements

This thesis is the result of five years of passionate work, struggle, happy moments and difficult times. These five years have been a great journey, where I learned many things that made me develop as a scientist and a man. I still eat gummy bears, though.

Thanks to Nelly, who was so brave to give me the chance to embark into the unexplored world of genomics and bioinformatics. You have been an example of kindness and scientific integrity, that I will never forget. Thanks to Brian, who has been patient enough to teach me the fundamentals of bioinformatics and programming. I thank all the members of the Pitteloud lab, amazing people who helped me so much during these years. Jim, pushing me to do always better. Thank you for your support and painstaking effort in getting this thesis done. Thanks to my Italian colleagues, Andrea and Sara, the *zii*. Your friendship has supported me throughout this time, and I am sure it will be the same in the future. Thanks to Yisrael, Hichem, Ginny, Jean-Pierre, Nadia, Andrew, Cheng, Celine, Estelle, Emmanuel, Caroline, Justine, Jenny, Niko.

My friends. The *piccola Beér* crew in Lausanne: the *bugnoni* Francesco and Floriana, Nicla, Michele, Giuliana, Maurizio, Enrica and Dante. I felt home having you around, and I will miss those moments together. Thanks to Andreas, Anne-Maud, Anastasia, Ali and Anand, as well as Beryl, Iris, Delphine. A special thanks to the Reymond and Pelet groups, who “adopted” me at conferences and lunches at Dorigny, respectively. My best friends in Italy, Lucia and Angelo. Seems like time has never passed for us, and after 20 years we are still here, supporting each other and laughing about what happens in our lives.

My Family in Italy. I will never be thankful enough for the gift of having you in my life. You always supported me during these years, starting from that famous day when I communicated on Skype that I had the PhD position. And right after we closed the call because everyone was crying. My *sparroni*, my parents, my “old” and “new” brothers and sisters, my uncles and cousins, my *bestiacce*, my *badidda*. You are a pillar of my life, I would never be the person I am without you.

My new Family. Ester, Luca, Sonia, Tania, Alice, Agnese, Ryan, Kifli, Thelma...thank you for having accepted me as part of the special Csikszentmihályi family. Love you all.

My All. Giulia, you are the miracle of my life. This thesis is all yours, without you I would never made it. You supported me with love, patience, advices, you backed me up during difficult periods and rejoiced with me in the happy moments. You gave me the self-confidence to go on, and I never gave up because you were always next to me. You are my sunshine.

Pietro Davide Cassatella, my son. You are a gift from Heaven. Since the first time your mother and I saw you, it looked like you were always part of our life. Despite the few hours of sleep in the first months, you gave me strength and many reasons not to give up. Your smiles always welcomed me when getting back home exhausted, and gave me energy to continue. I love you, we love you.

Last, but not least, this thesis is for my first niece, Gloria. I look forward to holding you in my arms. I don't know you yet, but I already love you with all my heart.

Isaiah 40:31

Abstract (English)

Congenital hypogonadotropic hypogonadism (CHH) is a rare disorder characterized by absent puberty and infertility due to GnRH deficiency. More than 30 genes have been implicated with the disorders, yet for 65% of CHH patients there is no genetic cause to explain their disease phenotype. To better characterize the genetic architecture of CHH and discover new genes associated with the disease, I performed whole-exome sequencing in a large cohort of CHH patients. I implemented a semi-automated bioinformatics pipeline to accurately process raw sequence reads and output high-quality variants for a series of genetic studies. Next, I investigated the genetic architecture of CHH by screening for the first time the largest set of known genes implicated in the disease. I observed that 51% of CHH patients have at least one putatively pathogenic mutation in a known gene, while 15% of patients have more than one known gene mutated (i.e., oligogenicity). Patients with congenital delay of growth and puberty, a common form of GnRH deficiency, do not show a genetic overlap with CHH. I then applied a biology-driven bioinformatics analysis targeting genes encoding fibronectin type-III domains proteins, which are critical for GnRH neuron migration. We identified loss-of-function mutations in two genes, *DCC* and its ligand *NTN1*, and demonstrated their role in GnRH neuron migration during early development in mouse and human. Last, I applied unbiased family- and population-based analyses to identify candidate genes associated with CHH. Interestingly, a *de novo* missense variant in the *SMC3* gene was detected in a patient with Kallmann syndrome (CHH and anosmia) and Cornelia de Lange syndrome (CdLS), a rare neurodevelopmental disorder. *SMC3* is a gene known to underlie CdLS and is part of the cohesin complex which is involved in

many cell processes, including transcriptional regulation. I then found additional CdLS patients diagnosed with KS or GnRH-related defects, suggestive of a phenotype-genotype overlap between CHH and CdLS.

Abstract (French)

L'Hypogonadisme Hypogonadotrope Congénital (HHC) est une maladie rare définie par une absence de puberté et une infertilité dues à un déficit en gonadolibérine (GnRH). A ce jour, plus de 30 gènes sont liés à la physiopathologie de l'HHC mais environ 65% des patients atteints ne présentent aucune mutation génétique connue. Afin de mieux comprendre la structure génétique de l'HHC et découvrir de nouveaux gènes impliqués dans cette pathologie, nous avons réalisé une analyse de séquençage complet de l'exome au sein d'une large cohorte de patients HHC. Dans un premier temps, nous avons développé un outil bio-informatique semi-automatisé permettant de discriminer les variants d'intérêt majeur à partir des résultats bruts de séquençage. Par la suite, nous nous sommes intéressés à l'architecture génétique de l'HHC en analysant tous les gènes connus de la maladie. 51% des patients présentent au moins une mutation pathogène dans un gène préalablement identifié du CHH, et dans 15% des cas, on observe la présence d'au moins deux gènes distincts mutés (i.e. oligogénisme). L'analyse génétique des patients présentant uniquement un retard pubertaire, forme la plus commune de déficience en gonadolibérine, n'a pas permis d'observer une architecture génétique similaire au HHC. Une approche expérimentale a complété notre stratégie bio-informatique en se focalisant sur les gènes codant pour des protéines à domaine Fibronectine-III, ces derniers étant particulièrement importants pour la migration des neurones à gonadolibérine. Dans ce sens, nous avons identifié et caractérisé des mutations "perte-de-fonction" au sein des gènes du récepteur *DCC* et de son ligand, *Netrin-1*. La dernière partie de notre étude porte sur l'identification de nouveaux gènes candidats potentiellement impliqués dans l'HHC à partir d'analyses familiale et de

population. Nous avons identifié une mutation spontanée (*de novo*) dans le gène *SMC3* chez un patient atteint conjointement d'un syndrome de Kallmann (HHC et anosmie) et d'un syndrome de Cornelia de Lange (CdLS, une maladie rare entraînant des altérations du développement cérébral). Le gène *SMC3* est connu pour être impliqué dans la physiopathologie du syndrome CdLS, et code pour une protéine du complexe de la cohésine qui contrôle différents processus cellulaires dont la régulation transcriptionnelle. Nous avons également identifié un hypogonadisme hypogonadotrope chez plusieurs patients CdLS, suggérant, pour la première fois, une association entre ces deux syndromes.

List of Abbreviations

ACMG/AMP: American College of Medical Genetics/Association of Medical Pathology
ALS: Amyotrophic lateral sclerosis
ANOS1: Anosmin 1
BAM: Binary Alignment/Map
BBS: Bardet-Biedl syndrome
BMI: Body Mass index
BP: Base pair
BQSR: Base quality score recalibration
BWA: Burrows-Wheeler Alignment
CDGP: Constitutional delay in growth and puberty
CdLS: Cornelia de Lange syndrome
ChAS: Chromosome Analysis Suite
CHD7: Chromodomain helicase DNA binding protein 7
CHH: Congenital hypogonadotropic hypogonadism
CNV: copy number variant
CoLaus: Cohorte Lausannoise
CTCF: CTCC-binding factor
dbSNP: database of Single Nucleotide Polymorphisms
DCC: Deleted in Colorectal Carcinoma
DIDA: Digenic Diseases Database
ExAC: Exome Aggregation Consortium
FGF8: Fibroblast growth factor
FGFR1: Fibroblast growth factor receptor 1
FISH: fluorescence in situ hybridization
FN3: Fibronectin type-III
FSH: Follicle-stimulating hormones
GATK: Genome Analysis Toolkit
GEMINI: GENome MINing
GnRH: Gonadotropin-releasing hormone
GNRHR: Gonadotropin-releasing hormone receptor
GW: Gestational week
GWAS: genome-wide association studies
HPG: Hypothalamic-pituitary-gonadal
HTS: High-throughput sequencing
IGSF10: Immunoglobulin superfamily member 10
IGV: Integrative Genomics Viewer
IL17RD: Interleukin 17 receptor D
InDel: insertions/deletions smaller than 50 base pairs
KISS1: Kisspeptin
KS: Kallmann syndrome
LDL: Low-density lipoprotein
LEP: Leptin
LEPR: Leptin Receptor

LH: Luteinizing hormone
MAF: Minor Allele Frequency
MGAT1: Mannosyl (alpha-1,3)-glycoprotein beta-1,2-N acetylglucosaminyltransferase
MRI: Magnetic Resonance Imaging
nCHH: normosmic congenital hypogonadotropic hypogonadism
NFE: non-Finnish European
NMD: Nonsense-mediated decay
NROB1: Nuclear receptor subfamily 0 group B member 1
NTN1: Netrin-1
OD: Osteoglophonic dysplasia
OMIM: Online Mendelian Inheritance in Man
PCR: Polymerase Chain Reaction
PCSK1: Proprotein convertase subtilisin/kexin type 1
PCSK9: Proprotein convertase subtilisin/kexin type 9
PFA: Paraformaldehyde
POLR3B: RNA polymerase III subunit B
PS: Pfeiffer syndrome
PTV: Protein-truncating variant
QC: Quality control
RS: Roberts syndrome
RVAS: Rare-variant association study
SA1: Stromal antigen 1
SAM: Sequence Alignment/Map
SBS: Sequencing-by-synthesis
SEMA3A: Semaphorin 3A
SIB: Swiss Institute of Bioinformatics
SLIT2: Slit guidance ligand 2
SMC3: structural maintenance of chromosomes 3
SNP: Single nucleotide polymorphism
SNV: Single nucleotide variant
SOX10: SRY-box 10
TAC3: Tachykinin 3
TAD: Topologically-associated domain
VNO: Vomeronasal organ
VQSR: variant quality score recalibration step
VUS: variant of unknown significance
WES: Whole-exome sequencing
WGS: Whole-genome sequencing
WS: Waardenburg syndrome
XLID: X-linked intellectual disability

Table of contents

Acknowledgements.....	4
1. Abstracts.....	6
2. Introduction.....	13
3.1. <u>Section 1</u> – Implementation of a bioinformatics pipeline to investigate the genetics of CHH using whole-exome sequencing.....	27
3.2. <u>Section 2</u> – Definition of the genetic architecture of GnRH deficiency disorders.....	42
3.3. <u>Section 3</u> – Biology-driven identification of new CHH genes.....	68
3.4. <u>Section 4</u> – New gene discovery with family- and population-based analyses.....	88
4. Thesis conclusions and future perspective.....	125
5. References.....	132
6. Supplementary material.....	149

Introduction

Understanding genetic causes of rare diseases: strategies and methods

Human genetics: from genomic to post-genomic era

Human genetics addresses a complicated and exciting challenge in modern science – discovering the molecular bases underlying phenotypes. Focusing on disease, many successful findings were achieved in the last decades linking rare DNA sequence changes (i.e., variants) to severe phenotypes. Genomic variants can range from large segments of the genome (e.g. duplications, deletions, translocations, inversions), down to insertions/deletions smaller than 50 base pairs (InDels). The smallest types of DNA sequence change are single nucleotide variants (SNVs) occurring in one of the 3.2 billion bases present in the human genome. Evaluating large-scale genomic changes, cytogenetics helped identifying disease-causing variations using karyotyping to detect chromosomal imbalances (e.g., trisomy of chromosome 21 in Down's syndrome¹), and fluorescence in situ hybridization (FISH) to identify inversions, translocations, copy number variants (CNVs), and complex chromosomal aberrations such as ring chromosomes or mobile elements insertions².

The methods used to identify a causative locus in a disease phenotype were mostly linkage studies and, thanks to Sanger sequencing³, positional cloning. This helped identifying the genetic causes of severe diseases such as Cystic Fibrosis⁴,

Duchenne Muscular Dystrophy⁵ or Type I Neurofibromatosis⁶. Sanger sequencing permitted to sequence an entire gene or some targeted areas of the genome of interest, often prior to supporting evidence linking the candidate gene to the disease (e.g., sequence homology with known genes, animal models⁷, role in key biological pathways, linkage studies⁸). However, the traditional paradigm of “one gene, one phenotype” demonstrated to be unfit in many circumstances. For example, each variant does not influence a phenotype in the same way. Kallmann syndrome (KS), Pfeiffer syndrome (PS), and Osteoglophonic dysplasia (OD) are three distinct diseases caused by heterozygous mutations in the *FGFR1* gene. The functional effects of *FGFR1* mutations on phenotypes, however, are different: loss-of-function *FGFR1* mutations underlie KS⁹, while gain-of-function mutations are implicated with PS^{10,11} and OD¹². Additionally, gain-of-function mutations in OD cluster within a highly-conserved region of *FGFR1*, encoding for extracellular juxtamembrane and transmembrane domains of the receptor^{12,13}. This is an example of the additional layer of complexity in genotype-phenotype correlations. The disadvantages of Sanger sequencing were primarily technical as well as on the study design. In fact, sequencing large genes or several regions in extended cohorts was a challenging task in terms of time and costs.

The Human Genome Project consortium and Celera Genomics, using two different technologies derived from Sanger sequencing (shotgun sequencing and paired-end sequencing), achieved in parallel the first major milestone of human genetics in 21st century: the full haploid reference human genome^{14,15}.

However, sequencing an entire genome required enormous efforts in terms of time and budget, and sequencing large cohorts to study common disease or non-

Mendelian disorders was impractical. Few years later, the release of the first haplotype map of the human genome, HapMap, provided for the first time information on one million single nucleotide polymorphisms (SNPs) in different populations¹⁶. Array technologies such as SNP arrays were critical to reach this goal, and were widely used in genome-wide association studies (GWAS). GWAS aim to identify correlations between common SNPs (MAF>5%) with traits or disease phenotypes, and have been the first large-scale genetic studies involving thousands of individuals. Since the very first GWAS study on 14,000 individuals with different common disease¹⁷, nearly 2,000 reports have described ~13,000 associations for many traits or diseases. GWAS led to the successful identification of SNPs that – although frequent in the population – exert strong effects for the manifestation of disease phenotypes. One of the most famous example is *PCSK9*, found to harbor common and low-frequency (MAF<5% and >1%) loss-of-function variants in individuals with low LDL-cholesterol^{18,19}. Despite the initial promising results, GWAS did not result as a major component for the discovery of new disease-causing loci. In fact, trait-associated SNPs in GWAS studies usually result – cumulatively – in small effects on the genetic heritability, with modest odds ratios (OR<1.5)²⁰, allowing to explain less than 10% of the genetic variation in disease and complex traits²¹.

Whole-exome sequencing as a powerful method to identify disease-causing loci

The advancements in sequencing technology have brought us to a new era of genetic research, allowing to generate sequencing data at high throughput, at high speed and low costs. High-throughput sequencing (HTS) allowed to directly sequence a

genome (or parts of it) in larger cohorts, shedding light on the genetic causes of many disease phenotypes.

HTS is the general definition of several massively sequencing approaches using non-Sanger technology. Sequences can derive from single DNA molecules of ~10 kb (e.g., PacBio SMRT) or from short reads of 75-300 bp (e.g., Illumina). In the case of short reads, a PCR amplification step is required to achieve a high-throughput set of sequences²². Sequencing-by-synthesis (SBS) technology was developed by Solexa (now Illumina) in 2008, and uses fluorescent dNTPs as reversible terminators, meaning that each dNTP needs to be removed after each round of sequencing. Before fluorophores are removed with an enzymatic reaction, a laser excites the molecules, producing four different signals depending on the bound nucleotide. This step is repeated at each incorporation of new fluorescent-terminated nucleotides by the DNA polymerase²³. Compared to Sanger, HTS is cheaper, faster and requires less DNA input to provide high-quality results.

One particular brand of SBS is whole-exome sequencing (WES). Briefly, WES targets the protein-coding regions of the genome (i.e., the exons of protein-coding genes), which account for only 1-2% of the genome (~30 Mb)²⁴. Prior to the steps described above, WES uses a hybridization step in which probes representing exonic sequences across the entire genome capture the target exons. Since its first successful use in Miller syndrome²⁵, WES has become a cost-effective methodology to study the genetic causes underlying Mendelian diseases. In fact, the exponential progress of HTS technologies is causing a substantial decrease in the costs of WES analysis²⁶, allowing researchers to screen larger cohorts of patients with high quality data (82% of genes

are covered for at least the 90% of bases called)²⁷. Although whole-genome sequencing (WGS) provides for a full mutational spectrum, WES is still considered as the first choice because of its more affordable costs and higher mean coverage²⁸. In relationship to Mendelian genetic disorders, WES is the most used technology as it has been estimated that 85% of the causative loci for Mendelian diseases reside in coding and functional regions of the genome^{29,30}. However, this estimate is likely biased, as exons of protein-coding genes have been the most prioritized regions in genetic studies. One major limitation of WES is the lack of a full mutational landscape that includes larger insertions/deletions (CNVs), chromosomal translocations/inversions, as well as the non-coding regions of the genome, critical for transcriptional regulation³¹. WGS has also been demonstrated to perform better than WES in uniform coverage and to provide higher quality variants in the exonic regions of the genome³². Thus, in order to describe the full picture of genomic variation, WGS is considered the most appropriate technology to be performed.

The technological advancements in DNA sequencing resulting from HTS now allow researchers to delve deeper into the complex genetic architecture underlying disease, involving cis- and trans-acting modifiers, epigenetic factors, and oligogenic inheritance (i.e. more than one mutated gene in the same individual). Specifically, oligogenicity has been observed in several heterogeneous genetic disorders such as Bardet-Biedl syndrome (BBS)³³, retinitis pigmentosa³⁴, amyotrophic lateral sclerosis (ALS)³⁵, syndromic Hirschprung's disease³⁶, and atrioventricular septal heart defect (AVSD)³⁷. The Digenic Diseases Database (DIDA)³⁸ currently collects information on 44 different diseases showing digenic inheritance, the simplest form of oligogenic

inheritance. In these instances, it is clear that oligogenicity helps to explain the variable phenotypes observed across and within families, previously explained with the *deus ex machina* concepts of incomplete penetrance and variable expressivity³⁹.

Congenital hypogonadotropic hypogonadism – clinical and genetic heterogeneity

CHH is a rare and complex form of GnRH deficiency

Fertility in mammals is controlled by the hypothalamic-pituitary-gonadal (HPG) axis and driven within the hypothalamus by a complex network of neurons secreting gonadotropin-releasing hormone (GnRH). Once secreted into the median eminence of the hypothalamus, GnRH reaches the anterior pituitary gland and triggers the release of luteinizing and follicle-stimulating hormones (LH and FSH) which in turn act in the gonads in order to control sex steroids synthesis and homeostasis (Figure 1)^{40,41}.

GnRH neurons display a very distinctive feature as, during embryonic development, they originate outside the brain, in the olfactory placode. Once fate-specified, GnRH neurons start their journey from the olfactory placode and then migrate along the olfactory axons crossing the cribriform plate, the olfactory bulb to reach the hypothalamus and start their maturation⁴² (Figure 1).

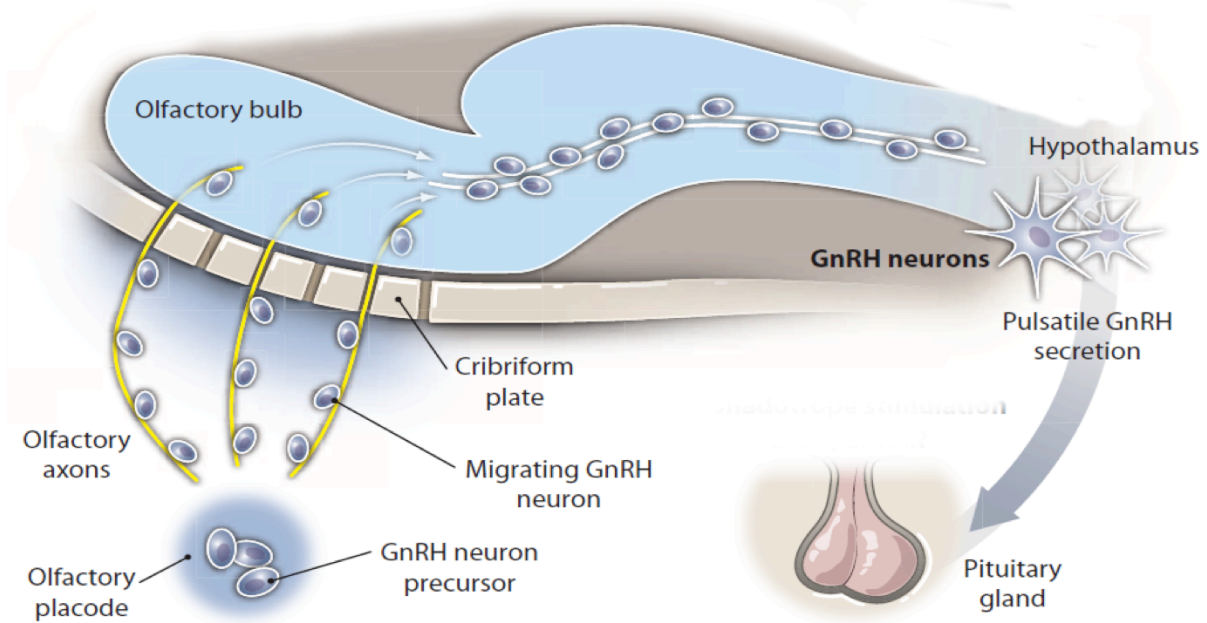


Figure 1. GnRH neuron ontogeny, migration and maturation processes. (From Sykiotis *et al.*, 2010)

Congenital hypogonadotropic hypogonadism (CHH) is a rare (1 in 4,000) congenital form of GnRH deficiency. A male predominance characterizes CHH, with a male to female ratio of 3-5:1^{43,44}, partially explained by X-linked recessive inheritance occurring only in men⁴⁵. Clinically, CHH patients typically present during adolescence with absent or incomplete puberty, leading to the lack of secondary sexual characteristics (e.g., absence of body hair, small genitalia, a lack of breast development and menstrual cycle in females), and infertility. Serum levels of sex hormones are low in CHH patients (testosterone <3 nmol/L, LH <2 UI/L, FSH <2 UI/L).

In approximately 50% of cases, CHH is associated with an absent or partial sense of smell (anosmia/hyposmia) and is termed Kallmann syndrome (KS). Patients with a normal sense of smell are considered to be normosmic CHH (nCHH). Other less frequently associated phenotypes are craniofacial (cleft lip/palate⁴⁴), skeletal

(craniosynostosis⁹, osteoporosis⁴⁶, split hand/foot malformations⁴⁷), neurological (synkinesia⁴⁸, hearing loss⁴⁹), or urogenital (bilateral/unilateral renal agenesis⁵⁰, cryptorchidism⁵¹) defects. Phenotypically, CHH exhibits variable expressivity and reduced penetrance both within and between families even when carrying the same mutation,⁵² however oligogenicity can partially explain this spectrum.

The genetic etiology of CHH: a complex and yet unsolved case

The first gene implicated in CHH was *ANOS1* (formerly *KAL1*), a gene coding for the protein Anosmin-1, an important factor for GnRH neuron migration during development. All male patients with mutations in *ANOS1* are diagnosed with KS, indicating an almost full penetrance of *ANOS1* mutations in hemizygous state⁵³. Various strategies have been used to discover new genes for CHH, such as linkage analyses in consanguineous families⁵⁴⁻⁵⁶, cytogenetic tests to identify chromosomal abnormalities^{9,57-59}, network analysis⁶⁰, candidate gene approach *via* Sanger sequencing⁶¹⁻⁷⁰ or, more recently, using WES^{54,71}. Mutations in many genes (n=33)^{56,72,73} have been directly linked to CHH (Figure 2).

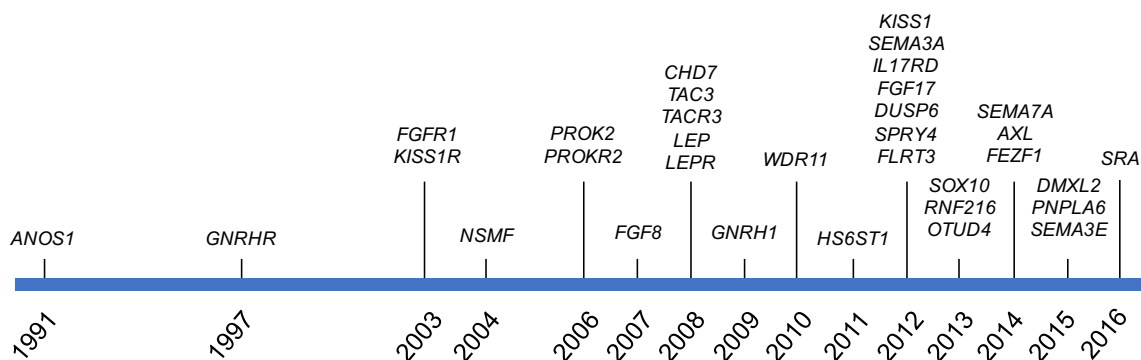


Figure 2. Timeline of discovery of genes implicated with CHH.

Notably, despite the successful genetic discoveries for CHH in last 20 years, only 30-35% of CHH patients have a clear genetic cause underlying their disease^{39,60}. Mutations in individual CHH genes have been demonstrated to be inherited in autosomal dominant (e.g. *FGFR1*⁹, *CHD7*⁶⁶, *SEMA3A*⁷⁴), autosomal recessive (e.g., *GNRHR*⁶³, *KISS1*⁵⁴, *TAC3*⁷⁵), and X-linked recessive (*ANOS1*⁵⁷, *NROB1*⁷⁶) patterns. *De novo* mutations consistent with autosomal dominant inheritance have also been observed for the *FGFR1*^{77,78}, *CHD7*⁷⁹, *SOX10*⁸⁰, and *FGF8*⁸¹ genes. However, these monogenic models of Mendelian inheritance do not completely explain the variable phenotypic expression both within and across families of the disorder. Recently, digenic and oligogenic inheritance has been demonstrated to at least partially explain such differences. Indeed, 2.5-7%^{39,60} of CHH patients harbor mutations in more than one gene, implicating a cumulative effect of known loci reportedly inherited with dominant, recessive or X-linked modes^{39,60}.

The genes underlying CHH participate in different stages and networks in GnRH biology. Genes involved in GnRH neuron fate specification⁶⁵, development and migration^{57,82} are more often associated with KS, while genes implicated in GnRH neuron homeostasis and maturation^{55,83} and GnRH secretion are primarily mutated in nCHH patients^{63,67}. However, KS and nCHH have only recently been considered as separate phenotypic subgroups of CHH. In parallel, a genetic overlap between KS and nCHH has also been shown, with several genes found to be mutated in both subgroups⁷².

Phenotypic and genetic overlap of CHH with other diseases

As described above, the multifaceted phenotypic features of CHH result in high genetic and clinical heterogeneity. The variable combinations of CHH with one or more associated phenotypes makes the genetic diagnosis of patients a challenging task. However, a thorough clustering of associated phenotypes in specific subgroups would enhance the diagnostic yield of genetic testing in such patients. While mutations in *ANOS1* are found in 5% of CHH patients overall, 43% of patients with CHH, anosmia and synkinesia have mutations in *ANOS1*⁸⁴. Only 2% of CHH patients are reported to have mutations in *SOX10*, however this increases to 35% in CHH patients with hearing loss⁶⁹. Similar frequency increases have been observed in CHH patients with hearing loss and *IL17RD*⁶⁰ or *CHD7*⁸⁴ mutations, or in CHH patients with split/hand foot malformations and *FGFR1*⁸⁵ mutations (Figure 3).

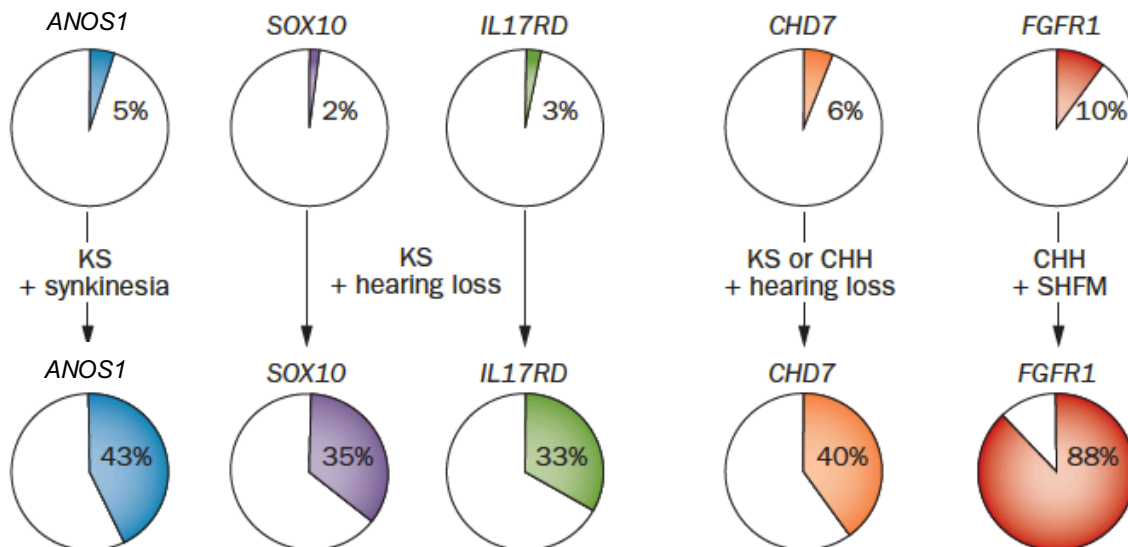


Figure 3. Associated phenotypes clustering in CHH increases the diagnostic yield in genetic testing.

Additionally, early onset of obesity and hypogonadism has been linked with mutations in *LEP*, *LEPR*⁸⁶ and *PCSK1*⁸⁷. Therefore, detailed phenotyping of CHH individuals is critical to enable successful gene discovery based on phenotypic clustering.

The defining phenotype of CHH, i.e. hypogonadotropic hypogonadism, has been described in >80 different diseases or syndromes, according to the Online Mendelian Inheritance in Man (OMIM) database (Table S1). Although these syndromes are primarily characterized by phenotypes not associated in CHH (e.g., mental retardation, cardiac defects, facial dysmorphism), the common feature of hypogonadotropic hypogonadism might represent a clue for a potential genetic overlap with CHH. The most common associated phenotypes in CHH (e.g., hearing loss, skeletal abnormalities, cleft lip/palate) are also identified in many other syndromes, thus resulting in a complex mixture of phenotypes that partially overlap with the clinical manifestation of CHH. This suggests that another type of phenotypic clustering of higher magnitude – targeting overlapping syndromes – may also represent a proxy for the identification of overlapping genetic architecture. There are few examples of how overlapping syndromes led to the discovery of important players in CHH genetic etiology, namely Waardenburg and CHARGE syndromes.

Waardenburg syndrome (WS) is characterized by hearing loss and pigmentation defects of the hair, skin and iris. It presents with varying severity and associated phenotypes, defined as WS types I to IV^{88,89}. WS type II and IV are caused by mutations in *SOX10*, a transcription factor and member of the SOX family, which plays an important role in neural crest and oligodendrocytes development⁹⁰. *SOX10* mutations in

WS are inherited in an autosomal dominant fashion, with most of them being protein-truncating variants (PTVs)^{91,92}. The presence of olfactory bulb agenesis – a hallmark feature of KS – is found in 88% of WS patients⁹³ led to the screening of the *SOX10* gene in KS patients. Heterozygous missense mutations were found in 8% of KS probands, with the majority of patients also presenting with hearing loss⁶⁹.

CHARGE syndrome (coloboma, heart defect, choanal atresia, retarded growth and development, genital abnormality, and ear malformation) is a multi-systemic disorder caused by mutations in *CHD7*⁹⁴, a large protein involved in chromatin organization⁹⁵, critical for neural crest cell migration during development⁹⁶. CHARGE patients have been reported to manifest CHH key phenotypes, namely hypogonadotropic hypogonadism⁹⁷, anosmia⁹⁸ and olfactory bulbs agenesis⁹⁹. The phenotypic overlap between CHARGE and CHH – specifically, KS – was then shown to be mirrored by a genetic overlap, as independent studies identified KS patients harboring disease-causing mutations in *CHD7*^{66,79,100}. Similar to the differences in types of *SOX10* mutations between WS and CHH probands, *CHD7* mutations in CHARGE patients are primarily PTVs, while heterozygous missense mutations are found mainly in KS patients¹⁰¹. The examples of Waardenburg and CHARGE syndromes highlight a potential genotype-phenotype correlation specifically related to the severity of the mutations identified. Hence, the identification and characterization of overlapping phenotypes in other diseases represents a successful strategy to identify new genes implicated in CHH.

Thesis aims

CHH is a rare disease characterized by a complex set of implicated biological processes and high clinical and genetic heterogeneity. This poses tremendous challenges to investigate CHH genetic architecture and use new models and technologies to discover novel CHH genes. In 2012, I joined Pr. Pitteloud's laboratory — a team which focuses on the clinical and genetic study of CHH. With my background in Medical Biotechnology and an eagerness to deepen my knowledge in human genetics, I embarked on a thesis in computational biology to investigate the genetic architecture and discover new genes involved in the pathogenesis of CHH using high-throughput sequencing techniques. Given the multi-disciplinary aspects of my thesis, my studies were co-mentored by Pr. Nelly Pitteloud (Chief of Endocrinology, Diabetes & Metabolism, CHUV) and Dr. Brian Stevenson (Senior Scientist at the Swiss Institute of Bioinformatics and Vital-IT). My PhD work has been focused to proposing different complementary approaches based on HTS and bioinformatics to study the genetic architecture of CHH.

The first objective was to implement a bioinformatics pipeline to investigate the genetics of congenital GnRH deficiency using whole-exome sequencing (WES) to secure high-quality data.

Next, using this pipeline, I have investigated the genetic architecture of CHH using WES focusing on known CHH loci. This step is critical in order to define the proportion of patients with known genetic causes, assess carefully oligogenicity – a

model that has been described for CHH – and observe specific genetic patterns that would guide the strategies for the identification of new disease-causing mutations. In this project, I was interested in assessing the genetic overlap with constitutional delay in puberty, a transient form of GnRH deficiency that is clinically indistinguishable from CHH in adolescence.

In order to facilitate the discovery of genes involved in CHH pathogenesis, I have then developed a biology-driven bioinformatics strategy using relevant information from the known CHH loci.

Finally, I have also explored unbiased approaches to identify novel CHH genes applying “traditional” and innovative genetic strategies. I have performed family-based analyses using WES coupled with array-CGH in CHH families to identify mutations inherited as *de novo*, recessive or X-linked. In this context, I have also tested the performance of rare-variant association studies (RVAS) to identify associated genes in CHH patients vs. controls.

Section 1

Implementation of a pipeline to investigate the genetics of CHH using whole-exome sequencing

Background and rationale

The high-throughput nature and the unprecedented possibility to screen all protein-coding genes brought by WES showed high success rates in genetic studies. Since the very first use of WES to find the genetic causes of Miller syndrome²⁵, hundreds of genes have been characterized in various rare diseases¹⁰², reaching estimated “exome sequencing solve rates” from 25%¹⁰³, 49%¹⁰⁴, up to 58%¹⁰⁵ of cases, depending on the disease type and their mode of inheritance (dominant, recessive or X-linked). However, publications highlight that targeted filtering processes need to be put in place in order to deal with the 20,000 to 50,000 variants identified in each sequenced exome^{105,106}. Prioritizing variants by means of quality, type (protein-truncating, nonsynonymous and splice-site), predicted deleterious effects and absence in the control population often reduces the number of putative disease-causing variants to 150-500 per proband¹⁰⁵. Based on the genetic architecture of the disease of interest and/or the cohorts studied, additional targeted strategies can further reduce the

number of putative disease-causing variants to a few, facilitating the discovery of loci implicated with the disease. However, before embarking on new gene discovery strategies, it is imperative to generate consistent and high-quality variants from the raw reads produced by DNA sequencing machines.

To do this, I implemented a semi-automated bioinformatics pipeline to process sequencing reads, perform quality control, and generate high-quality variants in a large cohort of individuals.

Methods

Informed consent and phenotyping

The subjects involved in this study provided written informed consent prior to study participation. Prof. Pitteloud and her clinical research team recorded phenotypic information. The diagnosis of CHH was determined by 1) absent or incomplete puberty by 17 years, 2) low/normal serum gonadotropin levels in the setting of low serum testosterone/estradiol levels, 3) otherwise normal anterior pituitary function, and 4) normal imaging of the hypothalamic-pituitary area¹⁰⁷. Olfaction was assessed by self-report and/or formal testing¹⁰⁸. When possible, family members were recruited for clinical and genetic studies.

CDGP patients were referred to specialist pediatric care in central or southern Finland for evaluation between 1982 and 2004. All patients met the diagnostic criteria for self-limited CDGP, defined as 1) onset of Tanner genital stage II two SDs later than population average (i.e. boys with testicular volume >3 ml after 13.5 years of age and in girls Tanner breast stage II after 13.0 years of age)¹⁰⁹. Medical history, clinical examination, and routine laboratory tests were performed to exclude chronic illnesses, and the diagnosis of CHH was ruled out by spontaneous pubertal development at follow-up. All patients were followed until near-full pubertal development was attained (i.e. Tanner stage 4).

The primary control cohort used was the “Cohorte Lausannoise” (CoLaus)¹¹⁰, as both phenotypic information on each individual and raw sequencing reads were

available. BAM alignment files from 197 European controls from the 1000 Genomes project¹¹¹, downloaded from the ftp site (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data>) were also utilized. Finally, variants from the 60,706 controls of the Exome Aggregation Consortium (ExAC)¹¹² (<http://exac.broadinstitute.org>) were also used for allele frequency data.

Cohort description

The CHH cohort consists of 183 probands, with a male to female ratio of 2.5:1, which is consistent with previously published reports⁷². The cohort includes 103 patients with KS (78 males, 25 females), 77 with normosmic CHH (50 males, 27 females) and 3 patients diagnosed during neonatal period (all males). The majority of the CHH probands are of European origin (86%, n=157), while the rest are of Asian (8%, n=14), Middle-Eastern (4%, n=7), mixed (2%, n=3) or African (1%, n=2) origins. Affected and unaffected family members (n=111) were recruited for the study when available. Familial inheritance, defined as having at least one additional family members with CHH, delayed puberty or defects in olfactory function, is present in 37% of cases (n=67). Sporadic cases comprise 28% (n=52) of the cohort, while family history is unavailable for the remaining 35% (n=64) probands. Familial cases included 22 trios (unaffected parents and affected proband), 10 quartets (a trio plus an affected or unaffected sibling), and 2 larger, complex families. In addition, there are 17 pairs of affected relatives (i.e. affected siblings or affected parent and child) that are used for the co-segregation analysis (see Chapter 4, table 6). The remaining cases in the cohort were *singletons*.

The CDGP cohort consists of 8 *singleton* probands collected by Prof. Pitteloud's clinical practice (all males, of European origin), and 64 *singleton* probands (56 males and 8 females, of Finnish origin) obtained through a collaboration with Prof. Dunkel (Barts & the London Medical School, London).

Based on the study design and the period when each analysis has been performed, a different number of subjects were studied (Table 1).

Table 1. Inclusion criteria and subjects studied in thesis sections.

Section (Analysis)	Inclusion criteria	Subjects included
1	CHH; cohorts 1-5; all ethnicities	n=183 (probands); n=111 (family members)
2	CHH; cohorts 1-4; European	n=116 (probands); n=77 (family members)
	CDGP; cohorts 1-4; European, Finnish	n=72 (probands)
3	CHH; cohorts 1-4; all ethnicities	n=133 (probands); n=85 (family members)
4 (Family-based)	CHH; cohorts 1-5; all ethnicities	n=183 (probands); n=111 (family members)
4 (Population-based)	CHH; cohorts 1-5; European	n=157

DNA preparation

Subjects' DNA was extracted from white blood cells using PureGene kit (QIAGEN). DNA quantity and quality was recorded using Nanodrop (ThermoFisher) and samples were sent to BGI in 5 separate cohorts between the years 2011-2016 to undergo whole-exome sequencing. As requested by the sequencing provider, DNA quantity was ≥ 3 μg per sample, and quality was measured by calculation of absorption via spectrophotometry. DNA was sheared into random fragments using the LE220 instrument (Covaris), hybridized with exome capture probes (Agilent SureSelect V2 for

first cohort, Agilent SureSelect V5 for all remaining cohorts), and then enriched *via* PCR reactions. Exome capture libraries were sequenced on the HiSeq2000 platform (Illumina), generating FASTQ paired-end files.

Variant calling and filtering

An overall schematic of the variant calling and filtering process is depicted in Figure 4. The computations were performed at the Vital-IT (<http://www.vital-it.ch>) Center for High-performance Computing of the SIB (Swiss Institute of Bioinformatics). Raw sequencing reads were mapped to the reference human genome (GRCh37) using the Burrows-Wheeler Alignment tool (BWA)¹¹³ version 0.5.7a to generate a SAM (Sequence Alignment/Map format) alignment file. During library preparation prior to sequencing, several PCR cycles are carried out in order to have a reasonable amount of DNA to sequence. However, this amplification step can introduce a bias due to the presence of read duplicates, which may lead to an overrepresentation of certain alleles that do not correspond to the real haplotypes. To avoid these biases, Picard (<http://broadinstitute.github.io/picard>) MarkDuplicates tool version 1.80 was used to mark PCR duplicates in the SAM alignment process. After the analysis, Picard MarkDuplicates generates a duplication metrics output and a compressed binary version of SAM alignment files called BAM (Binary Alignment/Map format). The RealignerTargetCreator from the Genome Analysis Toolkit (GATK)¹¹⁴ version 3.3.0 was used to identify noisy regions around InDels. A set of known reference InDels from dbSNP release 138 was integrated into the analysis to “train” the tool to recognize high quality InDels. IndelRealigner was used to perform the actual realignment around the

InDels to generate a realigned BAM file. GATK's BaseRecalibrator tool applied a base quality score recalibration (BQSR) on the BAM alignment file, taking into account multiple covariates such as sequencing chemistry, base position in the sequenced read, and sequence context. BaseRecalibrator used reference SNPs from dbSNP version 138 and known gold-standard InDels from 1000Genomes¹¹⁵ and Mills *et al.*¹¹⁶ datasets. Recalibrated BAM alignments were then used for variant calling by GATK's HaplotypeCaller in gVCF mode, an intermediate step that outputs sample-based gVCF files, that will be then jointly used for multi-sample variant calling. This step is recommended in large cohorts, as joint genotyping of many BAM alignment files requires unfeasible amounts of memory and computing time. Every variant is assigned with a Phred-scaled quality score that is calculated as $-10\log E$, where E is the error probability^{117,118}. This means that a variant with Phred-quality score of 10 has a 1 in 10 or 10% chance of being a false positive, while a Phred-quality score of 50 suggests an error probability of 1 in 100,000 or 0.001%, inversely 99.999% accuracy. SNVs and InDels were called together, using a Phred-scaled quality threshold for calling variants of 50. Any variant called with a Phred-scaled quality <50 was flagged as "low-quality". Merging all sample-based gVCF files resulted in a multi-sample raw VCF file. The raw VCF variant call file was then divided into a SNVs-only file and an InDels-only file to undergo a variant quality score recalibration step (VQSR). VariantRecalibrator tool was used to build the recalibration model, taking into account HapMap phase 3¹¹⁹ and 1000Genomes OMNI 2.5¹²⁰ variants as true positives for the SNVs set, and Mills *et al.*¹¹⁶ variants as true positives for the InDels set. In the final step, recalibrated SNVs and InDels were merged into a single multi-sample VCF.

Quality check on WES data

Prior to downstream analysis, quality control (QC) on the set of exome sequences was performed in order to evaluate for any possible technical bias. QC on the BAM alignments were carried out using Picard's CalculateHsMetrics, which outputs important metrics to estimate sample reads and alignment quality. The percentage of unique reads passing the vendor filter compared to the overall reads passing the vendor filter should be >85%, and the percentage of target bases reaching 10X coverage should be >90% (Dr. Keith Harshmann, Genomic Technology Facility, University of Lausanne, personal communication). PLINK/SEQ version 0.10 was used to run QC on the VCF variant calls, using the *i-stats* command to generate sample-based metrics including the total number of variants (optimal range: 20,000-50,000) and the ratio of transitions/transversions (optimal ratio: 3)¹²¹.

Variants annotation

Variants were annotated using SnpEff version 4.2 and dbNSFP version 3.2 to generate the following information: 1) allele frequency in 1000Genomes, ESP and ExAC controls, 2) predicted functional effects on missense variants using SIFT and PolyPhen-2, and 3) conservation scores using GERP++, PhyloP and PhastCons algorithms. Allele frequencies in CoLaus controls were directly calculated from the VCF file using an in-house script.

Results

Bioinformatics pipeline workflow

The workflow of the semi-automated bioinformatics pipeline to process WES data is illustrated in Figure 4, and detailed in the Methods section above.

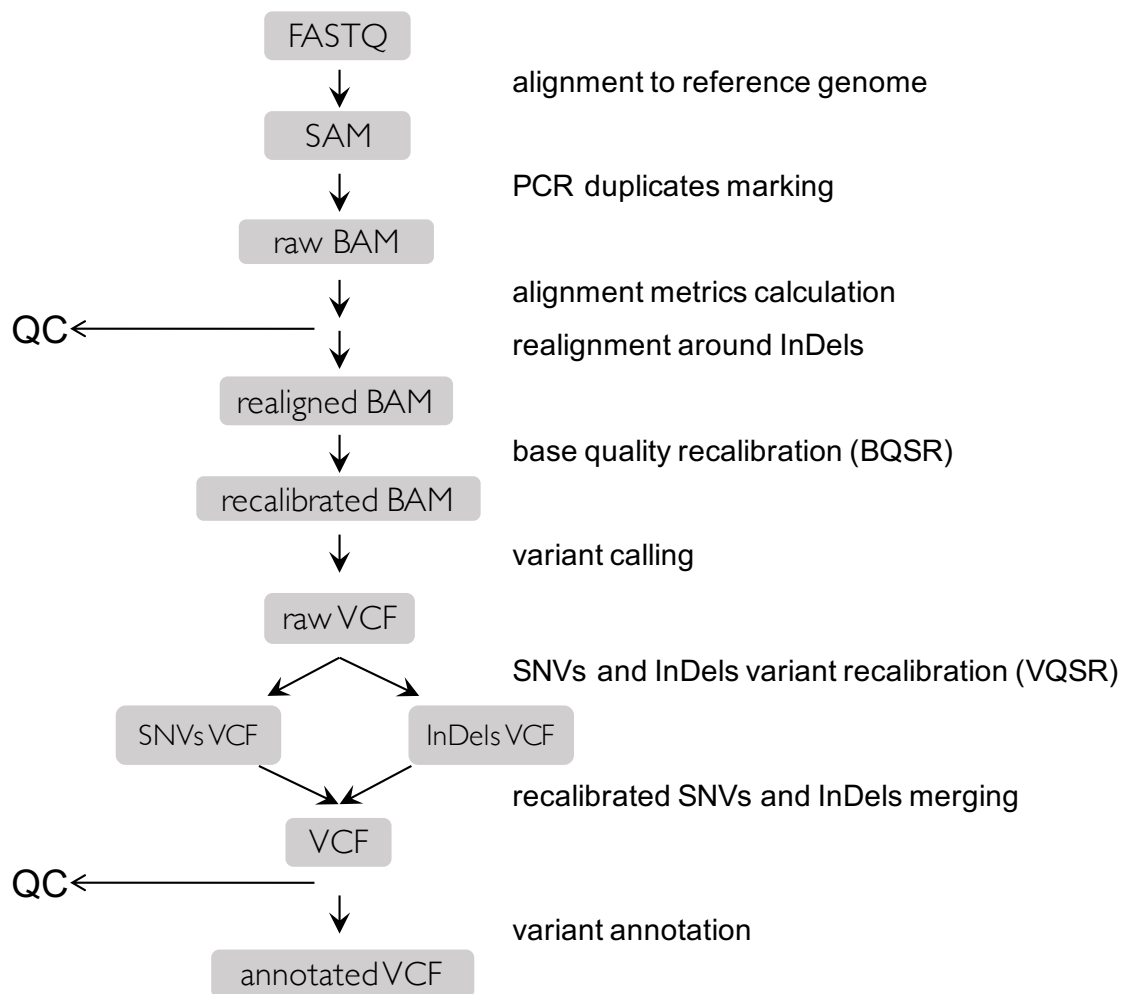


Figure 4. Workflow of semi-automated bioinformatics pipeline to analyze WES data in CHH patients.

Quality check on exome alignments

A low percentage of uniquely mapped reads suggests an overrepresentation of PCR duplicates, resulting in higher rate of false positives. There were no samples with low levels of uniquely mapped reads, and the majority of samples passed the stringent quality threshold of >85%. A nominal number of samples (8%) had slightly marginal quality for unique reads, and most of these were from the first cohort sent for sequencing (Figure 5).

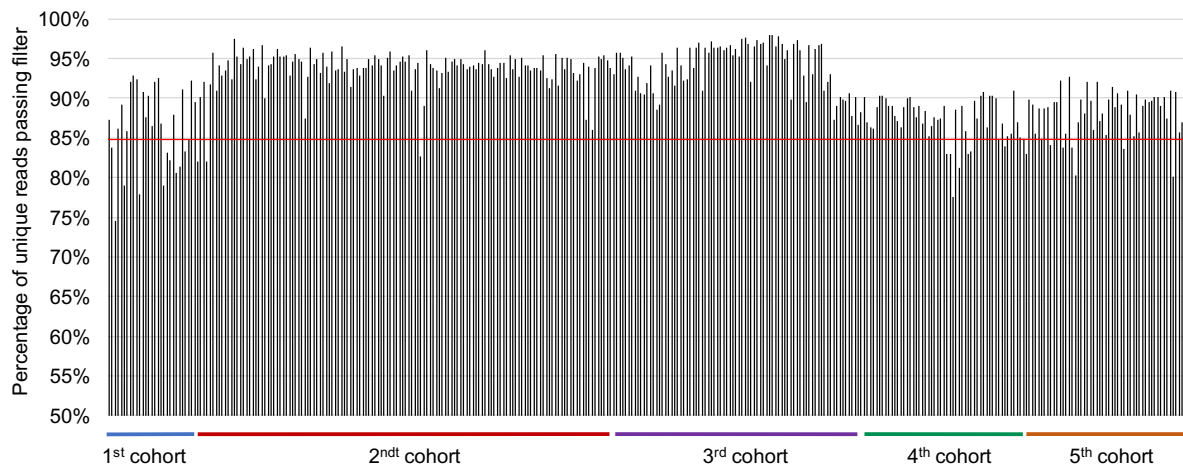


Figure 5. Percentage of unique reads passing filter in analyzed exomes. Percentage is calculated on the number of unique reads passing filter divided by the number of total reads passing filter. Samples are ordered in sequencing batches, from 1st (left) to 5th (right). Red line denotes good quality cutoff.

Additionally, a striking difference of the percentage of bases with coverage >10 reads (>10X) was observed in samples from first cohort relative to the remaining cohorts. However, all samples from first cohort were close to the passing quality threshold of 90% (Figure 6).

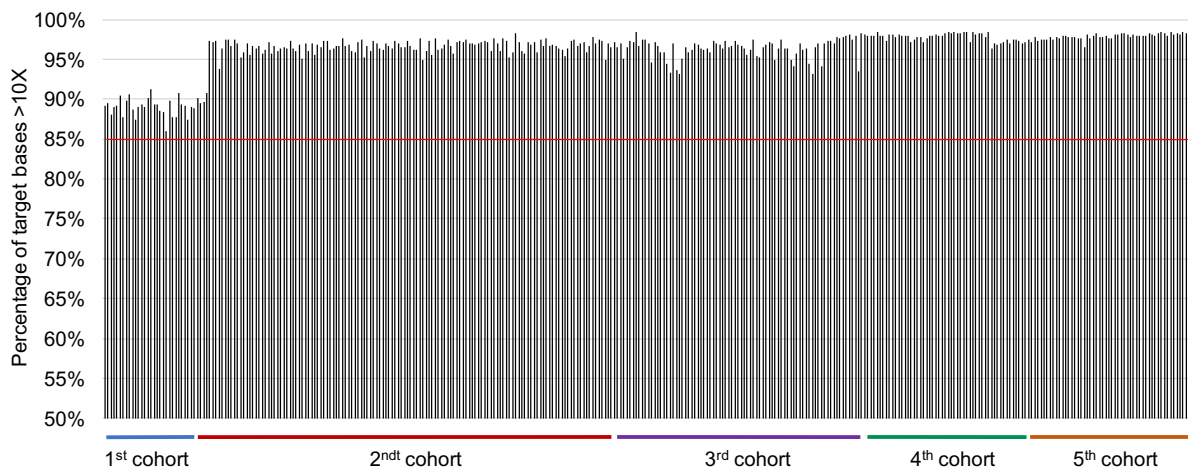


Figure 6. Percentage of target bases >10X in analyzed exomes. Samples are ordered in sequencing batches, from 1st (left) to 5th (right). Red line denotes good quality cutoff.

Quality check on variant calls

All samples showed transitions/transversions ratio close to 3, (median 2.96 ± 0.04), a cutoff typically used in WES data¹²¹. Similarly, all samples had a number of variants in the accepted range for a WES study (20,000-50,000). However, a lower number of variants in the first sequencing batch of 34 samples (median $23,936 \pm 555$ variants) was observed, while samples from the subsequent four cohorts had $\sim 30,000$ variants (median $30,245 \pm 1,059$).

Final list of variants

For subsequent downstream analyses, high quality ($GQ > 50$), rare ($MAF < 1\%$ in ExAC) coding and splicing variants ($n=92,793$) were retained (Table 2).

Table 2. Rare, high quality variants identified in sequencing cohort.

Variant type	# of variants
Missense	86,060
Frameshift insertion/deletion	2,666
Stop gained	1,576
Inframe insertion/deletion	1,190
Splice acceptor/donor	1,125
Start lost	125
Stop lost	51
TOTAL	92,793

Discussion

Bioinformatics analysis of HTS reads can be described in four main procedural tasks: 1) mapping reads to a reference genome, 2) calling variants, 3) annotating variants, and 4) analyzing variants according to the strategy of choice. Performing genetic analyses using high-throughput sequencing data seems a straightforward task, as many bioinformatics tools can process the large amount of sequences to correctly detect variants in the cohort of interest. However, many sources of errors can contribute to the introduction of bias in each of these steps. Similar to other scientific procedures, HTS can suffer of human errors during samples preparation, problems occurred during sequencing, or errors introduced by misuse of bioinformatics tools¹²².

In this project, bioinformatics processing of raw sequencing data down to high quality variants was performed using the Best Practices provided by the Genome Analysis Toolkit framework developed at the Broad Institute, USA¹²². The GATK Best Practices are very popular in studies that make use of HTS, providing high sensitivity and minimizing false positives. Other variant callers (e.g., SAMtools¹²³ and Freebayes¹²⁴) are also widely used in human genetics studies. Different benchmarking studies showed these algorithms to perform similarly, with minor differences in terms of precision and accuracy^{125,126}. However, the BWA-GATK HaplotypeCaller pipeline was used in this thesis for its robustness in calling both SNVs than InDels¹²⁵. It is notable that the same pipeline implemented in this thesis has been used to process 60,706 individuals in the Exome Aggregation Consortium, the largest public dataset of controls and widely considered as the reference to observe human variation¹¹².

During alignment and variant calling steps, quality control analyses were performed in order to identify any discrepancy at the cohort and sample level. Overall, the WES data show good to excellent quality. A small increase of PCR duplicates in <10% of the cohort was observed, and a sub-optimal number of bases reaching 10X coverage in all samples from the first sequencing cohort was noted. If a mutated nucleotide is covered by less than 10 reads, variant callers cannot estimate with certainty the presence of a variant, resulting in variants called with a low genotype quality¹²⁷. This might explain the lower number of high-quality variants in samples from first sequencing cohort compared with the other cohorts. The difference in terms of quality observed in samples from first cohort compared to others may also be due to the use of different exome capture probes used (Agilent SureSelect V2 for cohort 1, Agilent SureSelect V5 for all remaining cohorts). Agilent exome probes are widely considered to confer more robust and consistent results in WES studies, and the SureSelect V5 kit has been shown to yield a more complete coverage of genomic coding regions, higher read depth and higher percentage of bases on target¹²⁸.

Prior to joining the Pitteloud lab, no bioinformatics pipeline was set up to analyze HTS data in the research group. In this thesis, a semi-automated bioinformatics pipeline to process raw sequencing reads from whole-exome sequencing was implemented for a large cohort of CHH patients and family members, yielding a discrete number of high-quality variants that were used for the subsequent genetic studies of this thesis (Sections 2-4). To date, ~1,000 exomes have been processed with this bioinformatics pipeline, including individuals from other studies (CoLaus, 1000 Genomes) or for parallel projects in the Pitteloud lab. This has provided the foundation

of not only this thesis, but of multiple other projects completed or currently running within the department.

Section 2

Definition of the genetic architecture of rare and common GnRH deficiency

Background and rationale

Despite extensive effort by the scientific community, a full description of CHH genetic architecture is still missing. Little is known about a possible genetic overlap with other diseases characterized by transient GnRH deficiency as congenital delay of growth and puberty (CDGP).

Pubertal timing is a highly heritable trait with up to 50-80% of the variance explained by genetic factors¹²⁹. Congenital delay of growth and puberty (CDGP) and CHH are part of the same clinical spectrum, one being characterized by transient (CDGP) and the other by congenital GnRH deficiency (CHH)^{130,131}. Interestingly, a higher than expected proportion (10%) of family members of CHH probands report a history of delayed puberty¹³⁰, compared to 2.5% in the general population¹³². Reversal of hypogonadotropic hypogonadism in CHH patients after discontinuing hormone therapy also points to a clinical overlap between the two entities¹³³. In contrast to CHH, CDGP is a common disease and little is known about its genetic etiology. Currently, the differential diagnosis between CHH and CDGP during early adolescence remains

challenging, as both diseases present with isolated delay in puberty. Further, there are no specific biochemical markers to accurately differentiate these two disorders¹³⁴.

Genome-wide association studies (GWAS) evaluating common and rare variants in the timing of puberty identified significant associations with hundreds of loci, including regions near or within *ANOS1*, *TACR3*, *LEPR*, and *PCSK1* — 4 known CHH genes. Taken together, these loci account for <3% of the variance in age of puberty onset¹³⁵⁻¹³⁷. In view of the possible overlap between the pathophysiology of delayed puberty and conditions of GnRH deficiency, several studies have searched for variants in CHH genes in self-limited CDGP cohorts. A homozygous partial loss-of-function mutation in *GNRHR* was found in two brothers, one with CDGP and one with CHH¹³⁸, and an additional heterozygous mutation was found in one male with self-limited CDGP.¹³⁹ Of 50 CDGP patients investigated for sequence changes in *TAC3* and *TACR3*, only one variant in a single patient was found in the latter gene¹⁴⁰. However, a recent study screening 21 CHH genes in a CDGP cohort (n=56) found putative pathogenic mutations in 14% of patients¹³⁰. A recent study identified variants of intermediate frequency (MAF<2.5%) in *IGSF10* gene in 11% of CDGP probands¹⁴¹. *IGSF10* is a large protein that is part of the immunoglobulin superfamily, and appears to have a developmental role in GnRH neuron migration. In the same study, *IGSF10* was found to be mutated in 10% of CHH patients¹⁴¹, further supporting the hypothesis of a partial genetic overlap between CHH and CDGP.

In this perspective, I have used WES (and the bioinformatic pipeline established in Section 1) focusing on known CHH loci to investigate the genetic architecture of congenital GnRH deficiency (CHH) and assess the possible genetic overlap with CDGP.

Methods

Cohort description

The study cohort included 116 CHH probands of European descent (n=61 KS, n=55 nCHH) with a 2.5:1 male-to-female ratio consistent with previous reports of male predominance¹⁴². Within this cohort, 38% (n=44) of cases are familial (i.e. CHH, delayed puberty and/or anosmia in family members), while the remaining 72 cases (62%) are sporadic. The CDGP cohort consists of 72 unrelated probands with self-limited CDGP, and has been previously described in detail¹⁴³.

Definition of genes to be screened

CHH genes are defined as those which met the following criteria: 1) identified as CHH genes in “Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism – pathogenesis, diagnosis and treatment”⁷², 2) had publications demonstrating loss-of-function variants, 3) had been demonstrated to be expressed in organs/tissues relevant for GnRH biology, and 4) covered by the exome capture probes. Twenty-four genes met these qualifying criteria — *ANOS1*, *SEMA3A*, *FGF8*, *FGF17*, *SOX10*, *IL17RD*, *AXL*, *FGFR1*, *CHD7*, *HS6ST1*, *PCSK1*, *LEP*, *LEPR*, *FEZF1*, *NSMF*, *PROKR2*, *WDR11*, *PROK2*, *GNRH1*, *GNRHR*, *KISS1*, *KISS1R*, *TAC3*, and *TACR3*. In addition, the *IGSF10* gene was screened as it was recently shown to be implicated in CDGP and CHH¹⁴¹.

Variants filtering

Intronic variants within +/- 6 bp of exonic boundaries and predicted to affect splicing by MaxEnt¹⁴⁴ with a wild-type vs. mutated site change of +/- 20% were retained, as well as inframe/frameshift indels, stop-gain, and missense variants predicted to be damaging by SIFT¹⁴⁵ and/or PolyPhen-2¹⁴⁶ algorithms. Protein-truncating variants (PTVs) were defined as frameshift, stop-gain and splice variants, based on current consensus¹¹². For the purpose of this study, mutations are defined as 1) rare (MAF <1% in ExAC controls) PTVs, 2) rare missense variants predicted to be damaging to protein function by at least one *in silico* algorithm (SIFT or PolyPhen-2), and 3) loss-of-function variants based on *in vitro* studies, regardless of *in silico* predictions.

Statistical analyses

Statistics for individual and oligogenic variants in cases vs. controls were performed using a two-tailed Fisher's exact test.

Gene-collapsed rare variant association study (RVAS) was performed using a burden test which calculates the excess of rare alleles in cases compared to controls in a contingency table similar to a χ^2 test¹⁴⁷. Gene-based associations were calculated for CHH European probands vs. CoLaus controls, as well as for the KS and nCHH subgroups separately vs. CoLaus controls. Significant associations were then validated in a second, larger set of controls, using 33,370 non-Finnish European individuals from ExAC using a Fisher's exact test. Gene-based allele frequencies in ExAC were calculated dividing the sum of alternate allele counts in ethnically-matched samples by the average of the total alleles observed. Statistical significance in gene-collapsed RVAS was defined using

Bonferroni correction, dividing nominal significance (0.05) with the number of tests performed (i.e., genes analyzed, $n=25$); hence, the cutoff to determine significance was set at $p=0.002$.

Results

CHH genes are mutated in 51% of CHH probands

Exome sequences were analyzed on 116 CHH probands of European descent, and 59 (51%) harbored rare variants in 19 of the 24 CHH genes evaluated (Figure 7, Tables S2-3). No variants were identified in *NSMF*, *FEZF1*, *PCSK1*, *LEP* and *LEPR*. Nearly two thirds of familial CHH probands carried rare variants in CHH genes (27/44, 61%), while the frequency in sporadic probands was lower (32/72, 44%) (Figure S2).

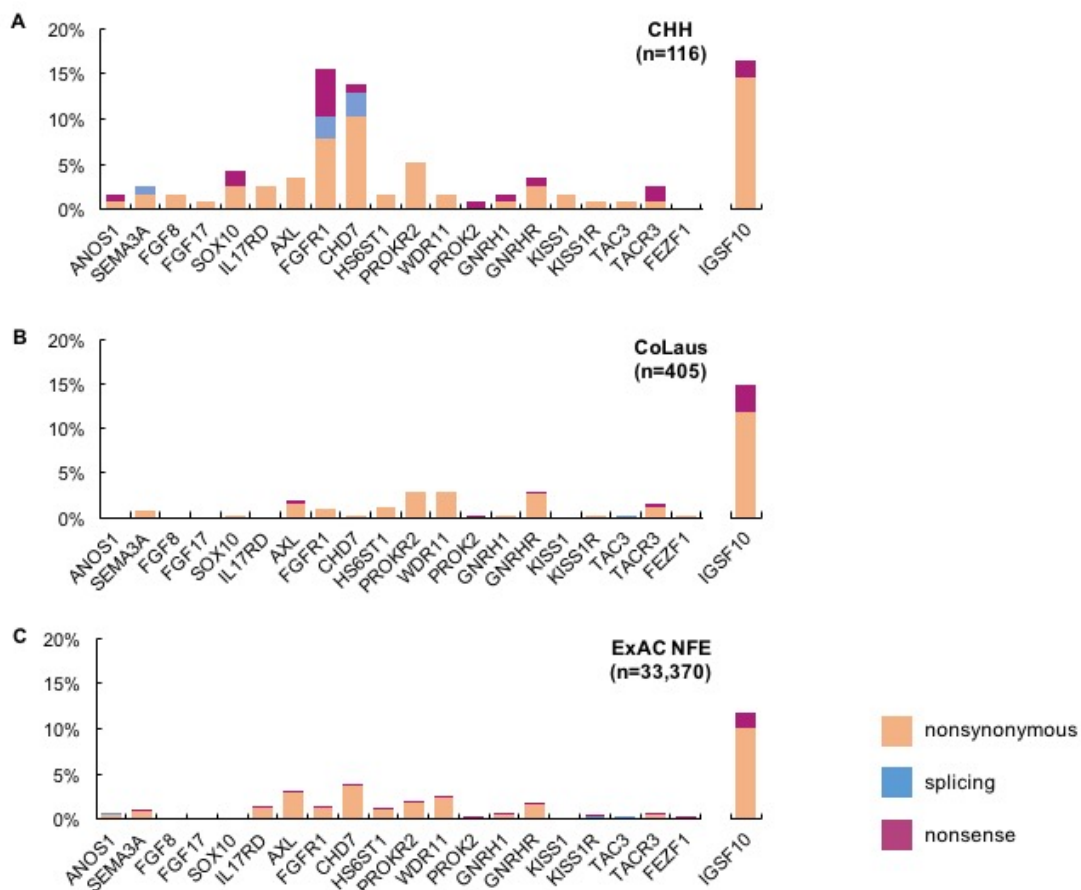


Figure 7. Mutation prevalence in CHH and control cohorts. Histograms showing CHH genes and *IGSF10* mutational prevalence in (A) CHH probands, (B) CoLaus, and (C) ExAC NFE controls.

FGFR1 and *CHD7* were the most frequently mutated genes in CHH probands (Figure 7), and both were statistically enriched compared to ExAC non-Finnish European (NFE) controls (Table 2, Figure S1). All of the identified *FGFR1* and *CHD7* variants were present in a heterozygous state (Table S2). In addition, a significant enrichment of rare variants was observed for *SOX10*, with a prevalence of 4% in CHH patients (Figure 7, Tables 3, S2). As previously described by our group in a different CHH cohort⁶⁰, an enrichment of rare variants has been confirmed in the FGF8 signaling pathway (*FGFR1*, *FGF8*, *IL17RD*, *FLRT3*, *HS6ST1*, *ANOS1*, *DUSP6*), present in 22% (25/116) of CHH probands.

Table 3. Percentage of mutated alleles in screened cohorts in CHH/CDGP known genes, and RVAS results in cases vs. controls.

Gene	CHH			KS			nCHH			CDGP *	CoLaus	ExAC NFE
	% mutated alleles	RVAS vs. CoLaus	RVAS vs. ExAC NFE	% mutated alleles	RVAS vs. CoLaus	RVAS vs. ExAC NFE	% mutated alleles	RVAS vs. CoLaus	RVAS vs. ExAC NFE	% mutated alleles	% mutated alleles	% mutated alleles
<i>ANOS1</i>	0.9%	ns	ns	1.6%	ns	ns	0.0%	ns	ns	0.0%	0.0%	0.6%
<i>SEMA3A</i>	1.3%	ns	ns	2.5%	ns	ns	0.0%	ns	ns	0.0%	0.4%	0.4%
<i>FGF8</i>	0.9%	ns	ns	1.6%	ns	0.002	0.0%	ns	ns	0.0%	0.0%	0.1%
<i>FGF17</i>	0.4%	ns	ns	0.8%	ns	ns	0.0%	ns	ns	0.0%	0.0%	0.1%
<i>SOX10</i>	2.2%	ns	4.4E-06	3.3%	2.8E-04	7.7E-06	0.9%	ns	ns	0.0%	0.1%	0.1%
<i>IL17RD</i>	1.3%	ns	ns	1.6%	ns	ns	0.9%	ns	ns	0.0%	0.0%	0.7%
<i>AXL</i>	1.7%	ns	ns	0.8%	ns	ns	2.7%	ns	ns	0.7%	1.0%	1.6%
<i>FGFR1</i>	7.8%	1.3E-08	8.5E-14	9.8%	1.8E-08	6.6E-11	5.5%	ns	1.1E-04	0.7%	0.5%	0.7%
<i>CHD7</i>	6.9%	1.6E-07	2.6E-05	9.0%	1.4E-09	4.2E-05	4.5%	1.2E-04	ns	0.0%	0.1%	2.1%
<i>HS6ST1</i>	0.9%	ns	ns	0.8%	ns	ns	0.9%	ns	ns	0.7%	0.7%	0.7%
<i>PROKR2</i>	3.0%	ns	ns	4.1%	ns	ns	1.8%	ns	ns	0.7%	1.5%	1.2%
<i>WDR11</i>	0.9%	ns	ns	0.8%	ns	ns	0.9%	ns	ns	0.0%	1.5%	1.2%
<i>PROK2</i>	0.9%	ns	ns	0.0%	ns	ns	1.8%	ns	ns	0.0%	0.1%	0.1%
<i>GNRH1</i>	1.3%	ns	ns	0.0%	ns	ns	2.7%	ns	ns	0.0%	0.1%	0.2%

<i>GNRHR</i>	3.0%	ns	ns	0.0%	ns	ns	6.4%	ns	5.7E-05	0.0%	1.5%	0.9%
<i>KISS1</i>	0.9%	ns	ns	0.0%	ns	ns	1.8%	ns	0.002	0.0%	0.0%	0.1%
<i>KISS1R</i>	0.4%	ns	ns	0.0%	ns	ns	0.9%	ns	ns	0.0%	0.1%	0.1%
<i>TAC3</i>	0.4%	ns	ns	0.0%	ns	ns	0.9%	ns	ns	0.7%	0.1%	0.1%
<i>TACR3</i>	2.2%	ns	ns	0.0%	ns	ns	4.5%	ns	2.1E-05	0.0%	0.7%	0.3%
<i>FEZF1</i>	0.0%	ns	ns	0.0%	ns	ns	0.0%	ns	ns	0.7%	0.1%	0.1%
<i>IGSF10</i>	9.5%	ns	ns	9.0%	ns	ns	10.0%	ns	ns	5.6%	8.1%	5.9%

ns: not significant; *: RVAS results in CDGP vs. controls is not shown, as no significant association was found.

Shared and specific genetic patterns in KS and nCHH

The mutational spectrum within the two subgroups of CHH (nCHH, n=55; KS, n=61) was examined. Among KS probands, *FGFR1* and *CHD7* were the most frequently mutated genes, and together with *SOX10* are significantly enriched when compared to controls (Figure 8A, Table 3). Notably, this enrichment was present when evaluating the whole CHH cohort, however it was more robust in the KS subgroup alone (Table 3). Interestingly, *FGF8* showed a prevalence of 1.6% in KS probands (vs. 0.1% in controls) yet this association was not evident in the CHH cohort as a whole (i.e. KS and nCHH together) (Table 3). Additionally, rare variants in *ANOS1*, *SEMA3A*, *FGF17* and *FGF8* were only found in KS probands.

Among normosmic (nCHH) probands, *FGFR1* and *CHD7* were also the most frequently mutated genes. Rare variants in *GNRHR* and *TACR3* were only found in nCHH probands (7% and 5%, respectively) (Figure 8B). In addition to an enrichment for variants in *FGFR1* in the nCHH, which was also seen in the KS and entire CHH cohort, a statistical enrichment was also observed for variants in *KISS1*, *GNRHR* and *TACR3* within the nCHH population vs. ExAC NFE controls. Interestingly, the enrichment in these genes was not observed in the KS and overall CHH cohort (Table 3).

In addition to *FGFR1* and *CHD7*, six other CHH genes (*SOX10*, *IL17RD*, *AXL*, *HS6ST1*, *PROKR2* and *WDR11*) were mutated in both KS and nCHH probands (Figure 8). This represents an increased genetic overlap in comparison to prior reports⁷². Overall, these results indicate both exclusive and shared genetic architectures for both KS and nCHH.

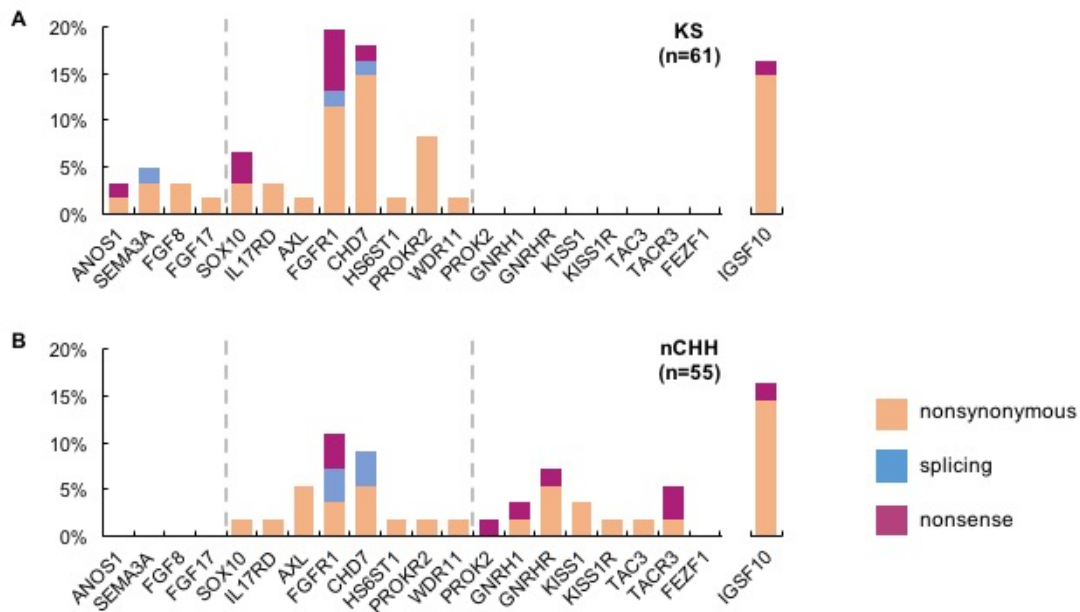


Figure 8. Mutation prevalence in KS and nCHH subgroups. Histograms showing CHH genes and *IGSF10* mutational prevalence in (A) KS and (B) nCHH probands.

Biallelic variants (i.e. homozygous or compound heterozygous in the same gene) were found exclusively in nCHH probands (11%), and were not seen in CoLaus ($p=2.3 \times 10^{-6}$), CDGP ($p=0.006$), or in KS clinical subgroup ($p=0.01$) (Figure 9).

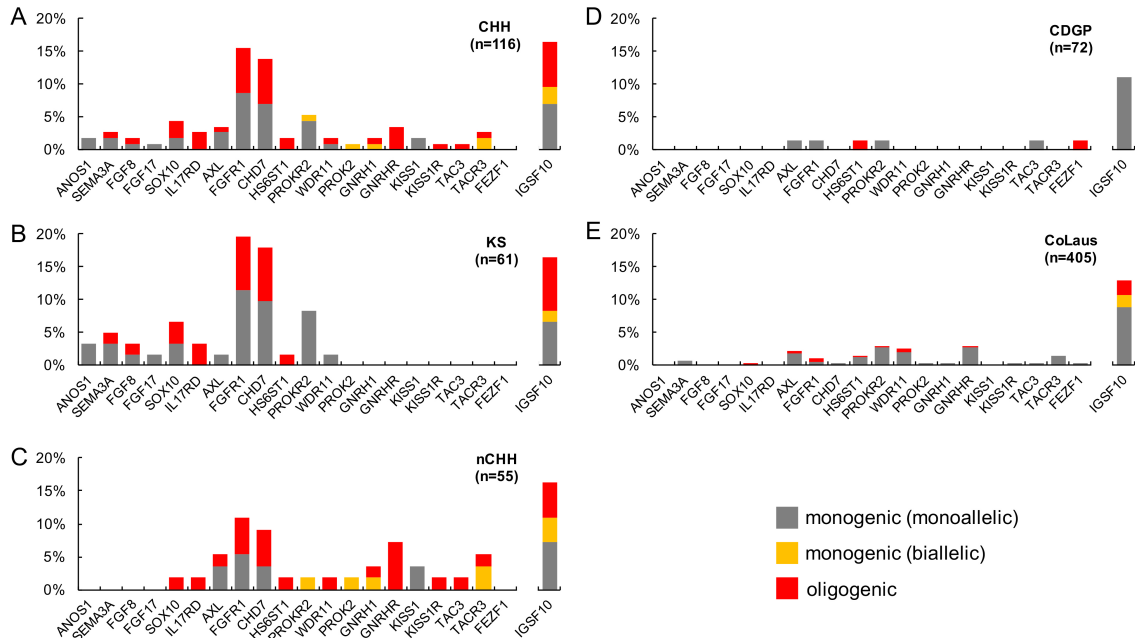


Figure 9. CHH genes mutations inheritance in screened cohorts. Histograms showing CHH/CDGP genes mutational prevalence in (A) CHH, (B) KS, (C) nCHH, (D) CDGP and (E) CoLaus screened individuals.

Furthermore, 27% (4/15) of genes mutated in nCHH (*GNRHR*, *GNRH1*, *PROKR2*, *PROK2*, *TACR3*) only exhibited biallelic variants, consistent with their recessive mode of inheritance (Figure 9C, Table S3). Notably, three nCHH probands harboring biallelic variants in *GNRHR* also carried additional heterozygous variants in *FGFR1*, *CHD7* and *WDR11*.

Limited genetic overlap between CHH and CDGP

Exome sequencing identified 7% (5/72) of CDGP probands harboring rare variants in CHH genes, all of which are present in a heterozygote state (Figure 10). Thus, the genetic profile of the CDGP cohort more closely resembles the controls (rather than CHH probands). Among the six identified mutations, there were five missense

variants and one intronic variant predicted to affect splicing. Three variants (*FEZF1* p.Ile337Lys, *FGFR1* c.190+6G>A, *TAC3* p.His83Pro) were not identified in 3,307 Finnish ExAC controls. Only one CDGP proband harbored two mutated genes (oligogenicity) (1.4%, $p=0.002$ vs. CHH), a similar rate as observed in controls (Figure 9D-E, Table S4).

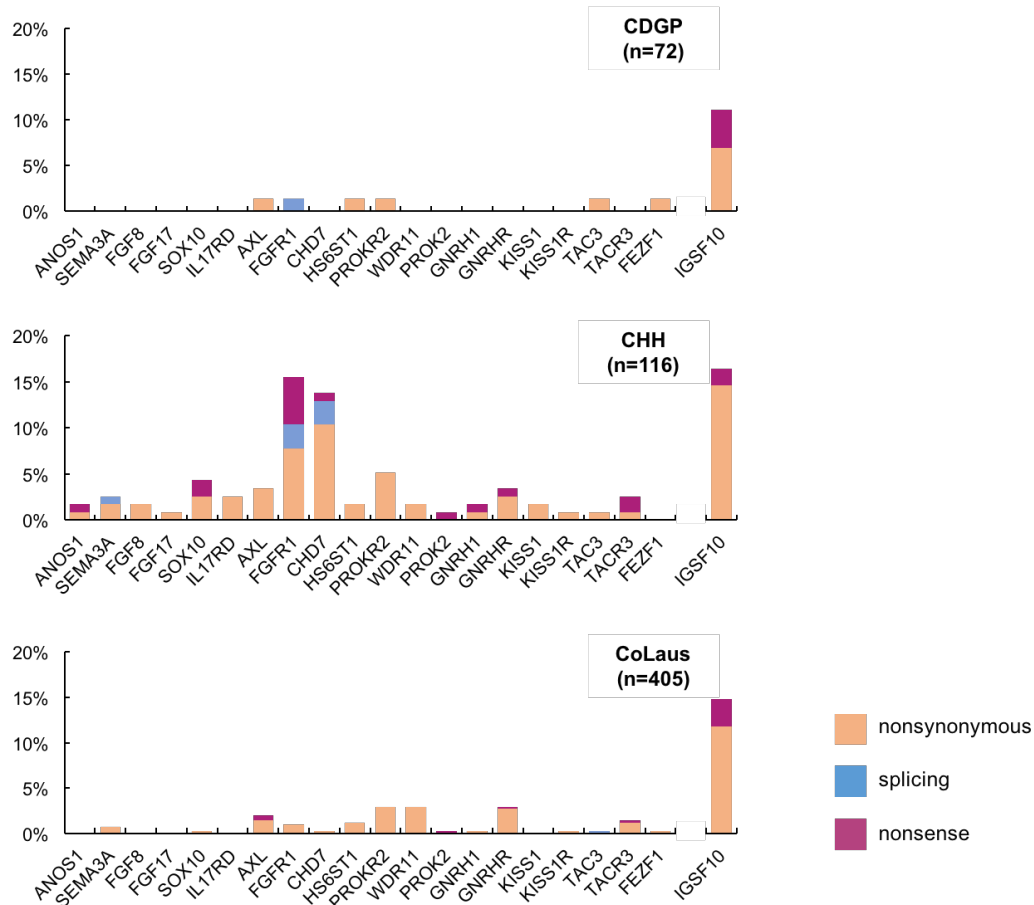


Figure 10. Mutation prevalence in CDGP, CHH and CoLaus individuals. Histograms showing CHH genes and *IGSF10* mutational prevalence in CDGP probands, compared to CHH and CoLaus (below).

Recently, mutations in *IGSF10* were identified in 10 CDGP families, segregating in an autosomal dominant fashion. The overall mutation rate for *IGSF10* was 13.2% in CDGP probands and 10.2% in CHH probands. Additionally, rare variant burden testing

showed an enrichment of *IGSF10* mutations in CDGP probands vs. controls¹⁴¹. In the present study, a similar frequency of putative *IGSF10* low-frequency variants has been found in CHH (14.7%) and CDGP (15.3%) probands using the criteria of Howard *et al.* (MAF <2.5% and deleterious predictions in both SIFT and PolyPhen-2)¹⁴¹. Notably, when applying the criteria used in this study a higher number of CHH patients (16.4%) harbored putative *IGSF10* mutations compared to CDGP (11.1%) (Figure 10).

Oligogenicity is a common factor in CHH patients

Oligogenic inheritance was present in 15% (17/116) of CHH probands (Figure 11) – a higher frequency than the 2.5-7% observed in previous reports^{39,60}. A significantly lower rate of oligogenicity was observed in CoLaus controls (2%, $p=6.4 \times 10^{-7}$).

Additionally, although monogenic inheritance was more frequent in KS (46%) compared to nCHH (25%, $p=0.03$), CDGP (6%, $p=3.7 \times 10^{-8}$) and CoLaus (16%, $p=4.6 \times 10^{-7}$), similar frequencies of oligogenicity were identified in both KS (13%) and nCHH (16%) (Figure 11, Table S4).

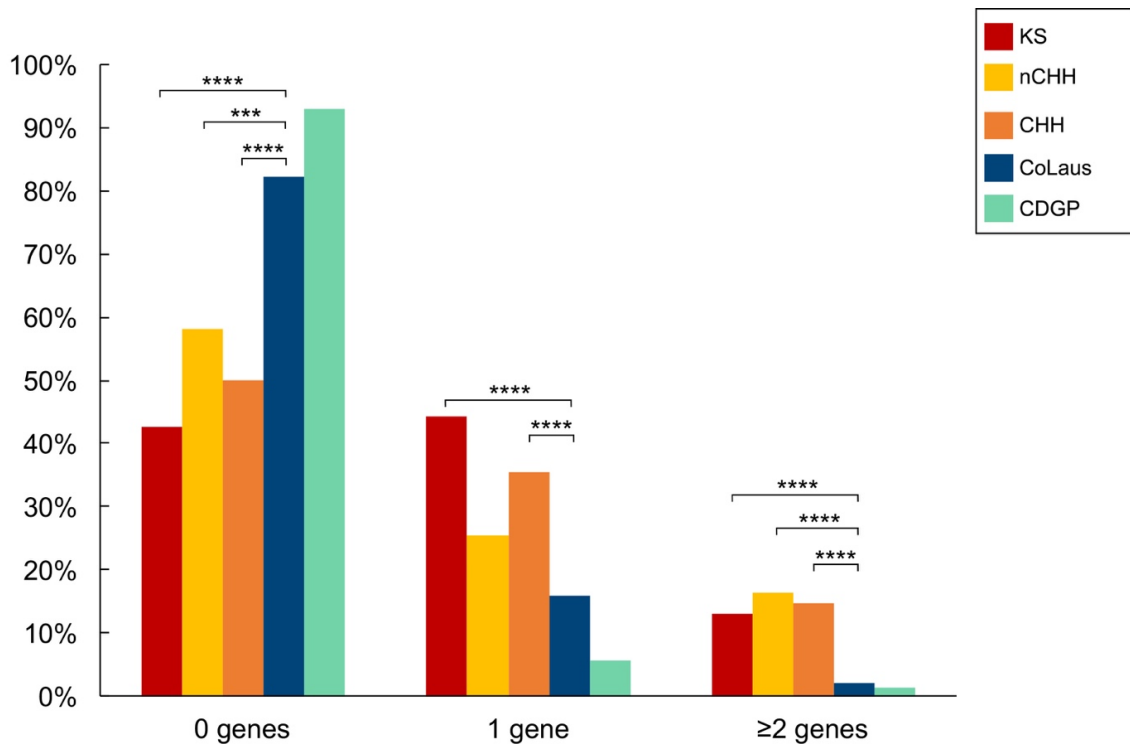


Figure 11. Oligogenicity patterns in CHH, CDGP and CoLaus. Histogram showing the frequency of KS, nCHH, CHH, CoLaus and CDGP individuals having no rare variants in CHH genes, one gene mutated or at least two genes mutated (oligogenicity). Differences between KS, nCHH and CHH vs. CoLaus controls were analyzed via a two-sided Fisher's exact test. $p < 0.05$ was considered significant. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. Not significant differences among KS, nCHH and CHH vs. CoLaus are not displayed.

Among possible gene combinations, *FGFR1* with *CHD7* was the most frequent pair observed ($n=4$), followed by *FGFR1/IL17RD* and *CHD7/HS6ST1* ($n=2$) (Figure 12A). When available, exome sequencing was performed on the family members of CHH probands displaying oligogenic inheritance. One KS proband (Figure 12B, Pedigree 1) carrying both *CHD7* (p.Pro369Ala) and *FGFR1* (c.1430+1delG) variants had two daughters after receiving fertility treatment. One of them carried both variants and was eventually diagnosed with KS, while the second daughter was unaffected and did not harbor either of the two changes. In Pedigree 2, three mutated genes (*FGFR1* p.Tyr654X, *CHD7* p.Leu2806Val and *SOX10* p.Asp64Val) were identified in a KS proband. Both

parents were deceased, however his KS brother could be screened and was shown to harbor the same *FGFR1* and *SOX10* mutations. There were no phenotypic differences between the proband and his sibling, therefore the absence of the *CHD7* variant in the affected brother likely rules out *CHD7* in the etiology of KS for this pedigree. Last, a KS proband was identified (Pedigree 3) showing a digenic inheritance of an *IL17RD* mutation (p.Tyr379Cys) from his anosmic mother, and a *de novo* *FGFR1* variant (p.Gly348Arg). No rare variants in CHH genes were detected in the anosmic father or the unaffected brother.

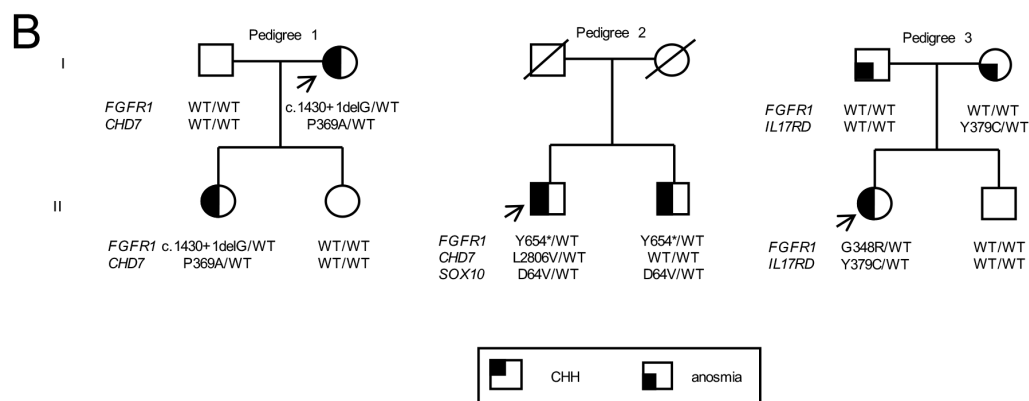
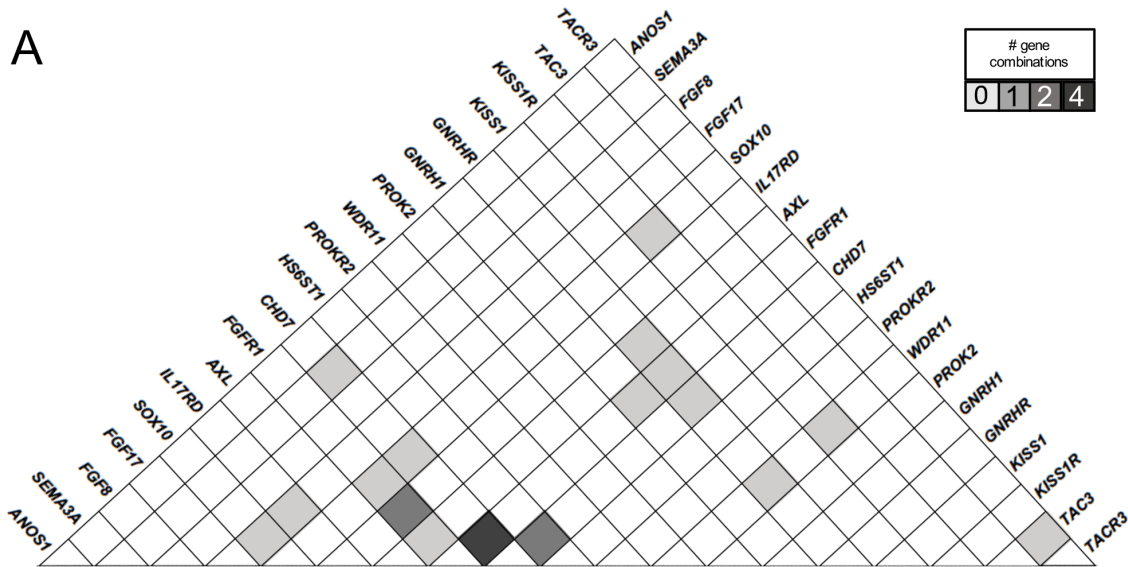


Figure 12. Most frequent oligogenic combinations and examples in pedigrees. (A) Matrix showing color-coded frequencies CHH genes digenic combinations. (B) Available pedigrees of CHH probands displaying oligogenic inheritance. Circles denote females; squares denote males; arrows depict probands; WT denotes wild-type; a diagonal slash indicates subject is deceased.

When assessing the rare variants identified in CHH probands, more than half (56%, 38/68) were not found in the ExAC NFE controls, and therefore are private. Also, a larger number of these variants (87%, 59/68) were not seen in the CoLaus controls either, however this larger frequency may reflect the smaller sample size of CoLaus. The presence of three particular variants in controls (*PROKR2* p.Leu173Arg, *GNRHR*

p.Gln106Arg and p.Arg206Gln) is not unexpected, as they are known founder mutations¹⁴⁸.

In total, 1,492 rare variants were observed in ExAC non-Finnish European controls (n=33,370), 80 variants in 18% of CoLaus controls (72/405), and 35 variants in 16% of 1000 Genomes European individuals (32/197). However, the majority of variants in CoLaus (89%) (Figure 9E) and 1000 Genomes European controls (90%) occurred in a monoallelic pattern. Given the variant-based (rather than sample-based) nature of the ExAC database, the allelic patterns in these controls could not be assessed.

Protein-truncating variants are enriched in CHH probands

Protein-truncating variants (PTVs) are defined as changes predicted to severely disrupt genes, i.e. splicing, frameshift and stop gain variants. A large fraction of the variants observed in CHH probands were PTVs (29%) while the frequency was significantly lower (5%) in ExAC NFE controls ($p=1.0 \times 10^{-9}$). Overall, 18% (n=21) of patients in our cohort harbored at least one PTV in the known CHH genes. Specifically, the CHH cohort was enriched for splice variants in *FGFR1* (2.6%, $p=1.7 \times 10^{-4}$), and for frameshift/stop gain variants in *FGFR1* (8%, $p=6.9 \times 10^{-13}$), *SOX10* (1.7%, $p=1.2 \times 10^{-5}$), and *TACR3* (1.7%, $p=4.9 \times 10^{-3}$) when compared to ExAC NFE.

A recent study evaluated the probability of a gene being intolerant to PTVs (probability of being loss-of-function intolerant, pLi) as a method to explain neutral and disease-causing PTVs seen in the control population¹¹². Genes with pLi scores ≥ 0.9 are defined as intolerant, thus PTVs in these genes are more likely to be pathogenic. Within the CHH cohort, 80% of PTVs were in intolerant genes, which is significantly higher

than ExAC controls (22%, $p=0.002$). Conversely, the majority of PTVs in controls (60%) were present in highly tolerant genes ($pLi \leq 0.1$) (Figure 13).

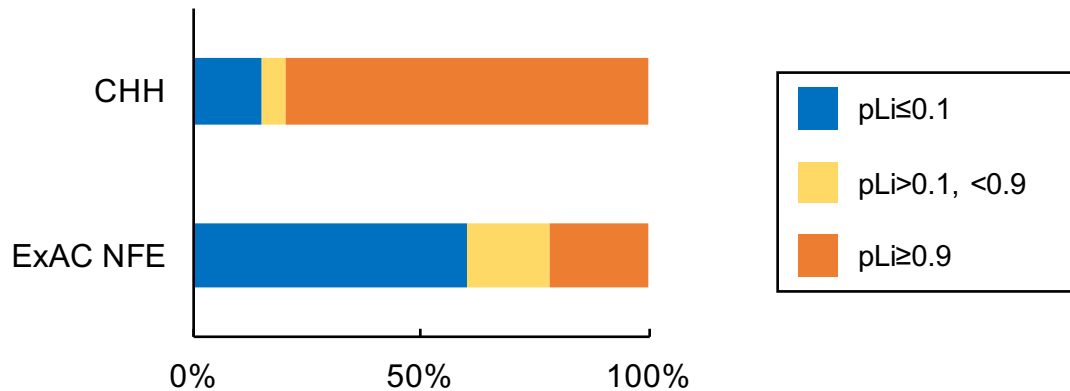


Figure 13. CHH patients are enriched with LoF mutations in LoF-intolerant genes.

Frequency of LoF mutations occurring in LoF-intolerant (orange) and in LoF-tolerant genes (blue) in CHH and ExAC NFE. Intermediate pLi scores are shown in yellow.

Interestingly, the vast majority of CHH genes showing to be tolerant to PTVs are inherited in an autosomal recessive mode, suggesting the higher number of PTVs than expected in controls is due by the high frequency of asymptomatic heterozygous carriers in the controls population. Similarly, PTV-intolerant CHH genes are inherited in autosomal dominant mode, implicating a heterozygous PTV might be sufficient to cause the disease phenotype (Table 4).

Table 4. CHH known genes and tolerance scores to PTVs.

	Gene	Inheritance	pLi
PTV tolerant	<i>IL17RD</i>	AD/oligogenic	0
	<i>TACR3</i>	AR	0
	<i>KISS1R</i>	AR	0
	<i>PROKR2</i>	AR	0
	<i>GNRHR</i>	AR	0.01
	<i>TAC3</i>	AR	0.04
	<i>NSMF</i>	AD	0.05
	<i>WDR11</i>	AD	0.07
	<i>PCSK1</i>	AR	0.09
	<i>GNRH1</i>	AR	0.1
	PTV intolerant	<i>FGF17</i>	AD
<i>HS6ST1</i>		AD	0.91
<i>SOX10</i>		AD	0.91
<i>FGF8</i>		AD	0.93
<i>ANOS1</i>		XLR	0.94
<i>SEMA3A</i>		AD	0.99
<i>FGFR1</i>		AD	0.99
<i>CHD7</i>		AD	1
<i>LEPR</i>		AR	1

AR: autosomal recessive; AD: autosomal dominant; XLR: X-linked recessive. *PROK2*, *NROB1*, *FEZF1*, *KISS1*, *AXL* genes are not shown as their pLI scores are between 0.1 and 0.9.

Furthermore, 80% of PTVs found in CHH probands (16/20) are heterozygous and occur in CHH genes traditionally defined as being inherited in an autosomal dominant mode (Table S5). All of the PTVs identified in CHH probands likely result in haploinsufficiency, as they do not lie within the last exon (or in last 50 bp of penultimate exon) and are therefore likely to be subjected to nonsense-mediated decay (NMD)^{149,150}.

Alternative methods of defining mutations do not affect results

In order to assess oligogenic inheritance, a MAF threshold of 1% was used, together with widely-used algorithms to predict damaging effects to protein function. To test whether these criteria were stringent enough, our data were re-analyzed with more rigorous criteria for minor allele frequency (<0.1%) and *in silico* functional

predictions for missense variants (deleterious prediction by both SIFT and PolyPhen-2). Overall, only two genes showed differences in the frequency or prevalence of mutations in CHH probands using these revised selection criteria — *CHD7* and *PROKR2* (Figure S3). When a more stringent definition of missense mutations was used the prevalence of *CHD7* mutations changed from 14% to 9%, and the prevalence of *PROKR2* changed from 6% to 3%. The reduction in *PROKR2* prevalence was due to the removal of the p.L173R founder mutation, present in four CHH probands and already demonstrated *in vitro* to be loss-of-function¹⁵¹. The remaining CHH genes were not significantly impacted by more stringent protein prediction or MAF criteria.

Identification of known CHH genes variants using family-based analyses

Seven known CHH genes gave positive results using the *de novo*, autosomal recessive and X-linked inheritance modes, demonstrating proof-of-principle for these strategies. The most frequently identified CHH gene in this analysis was *FGFR1*, harboring three *de novo* mutations in two KS patients and one with neonatal diagnosis of CHH. Interestingly, the two KS unrelated probands harbored a *de novo* variant occurring at the same nucleotide of the exon 9 (c.1135G>A, p.Gly379Arg), suggestive of the presence of a hotspot. One of the *de novo* mutations was a potential mosaic, as suggested by its alternate allele ratio of 26% in WES and the uneven overlapping peaks in Sanger sequencing (Figure S4). Targeted deep amplicon sequencing using >5,000X coverage identified a similar alternate allele ratio (25.2%) (data not shown), further supporting the mosaic nature of this *de novo* variant. A heterozygous *de novo* *CHD7* mutation (p.Trp908Arg) was also observed in a KS patient with CHARGE-like

phenotypes, presenting with hearing loss and external ear anomalies. This variant has never been observed in any controls database.

A *TACR3* stop gain homozygous variant (p.Trp275X) in two affected siblings diagnosed with nCHH, while both carrier parents had a history of delayed puberty but no consanguinity. This specific variant has been identified in nCHH patients in previous reports^{75,152}, and has been demonstrated to lie within a founder haplotype in Caucasian individuals¹⁴⁸.

Lastly a private *ANOS1* stop gained variant (p.Tyr630X) in a sporadic KS patient was observed, and was inherited from his asymptomatic mother.

Given the highly penetrant nature of the above genes, families containing these mutations were not utilized for subsequent family-based analyses for new gene discovery.

Discussion

CDGP and CHH are part of a continuum of GnRH deficiency, ranging from transiently delayed to a complete absence of puberty. However, it is challenging to make a distinction between CHH and CDGP in adolescent patients presenting with delayed puberty. In this project, the genetic overlap between these two conditions was investigated by focusing on rare variants in known CHH genes and *IGSF10*, a gene recently identified in CDGP. CHH and CDGP differ in terms of the number of patients harboring mutations in individual CHH genes (51% vs. 7%), as well as the overall mutational load (oligogenicity). In both instances, the CDGP probands more closely resembled the control cohort. Similar frequencies of *IGSF10* mutations in CDGP and CHH probands was observed, although higher than previously reported¹⁴¹. However, the association of *IGSF10* mutations in CDGP or CHH relative to controls was not replicated.

Recent GWAS studies have identified hundreds of loci associated with age at menarche in females^{135,137} and voice breaking in males¹³⁶ in the general population, several lying close to or within CHH genes. These data suggest a genetic overlap between CHH and CDGP. However, GWAS signals may result from intergenic, intronic, promoters, or other regulatory changes that are not detected by exome sequencing. Therefore, our results in CHH and CDGP patients would therefore have missed pathogenic mutations in regulatory regions. Notably, a genome-wide significant signal in the coding sequence of *TACR3* (p.Trp275X) was reported in nCHH probands in this report as well as in previous studies^{60,75,153}. Although prior GWAS studies have not

identified an association for its ligand *TAC3*, mutations in *TAC3* in both CHH and CDGP cohorts were found in the current study. Further, *TAC3* mutations were previously reported in CHH as well as CDGP¹³⁰. *TAC3* gene codes for neurokinin B, a modulator of GnRH secretion expressed in the hypothalamus together with kisspeptin and dynorphin¹⁵⁴. Combined, these data implicate the neurokinin B pathway in both CHH and CDGP. Larger studies examining pathways rather than individual genes identified by GWAS are likely required to further elucidate the genetic overlap between CHH and CDGP.

Using HTS to examine a larger number of CHH genes in our study, mutations were found in 51% of CHH cases. This is increased in relation to prior reports of 31%³⁹ and 35%⁶⁰ respectively. However, roughly half of CHH probands still do not have an identified genetic etiology. Oligogenic inheritance is present in 15% of cases of our cohort, a higher frequency as compared to 2.5% and 7% shown in previous reports^{39,60}. These results might be in part due to the increased number of CHH genes screened in our study.

A genetic overlap between KS and nCHH was identified, as well as specific and distinct pattern of mutated genes. Using gene-based rare variant association studies (RVAS) on the entire CHH cohort, significant associations for *FGFR1*, *CHD7*, and *SOX10* were found. Notably, *FGFR1* and *CHD7* are the most frequently mutated genes accounting for mutations in ~30% of the entire CHH cohort, while *SOX10* was mutated in 4.3% of probands. After categorizing CHH for KS and nCHH, the gene-collapsed associations remained significant for *FGFR1* in both subgroups while *CHD7* and *SOX10* were significant only for KS. Notably, a significant association appears for *FGF8* in KS,

while *GNRHR*, *TACR3*, *KISS1* showed association only in nCHH. However, it is important to stress that RVAS cannot represent the surrogate of functional and more extended genetic studies. Unidirectional burden tests will not be able to find associations in loci showing low penetrance and variable expressivity - even with larger cohorts, as they will be mutated at remarkable frequencies in controls as well. An example of this is *KISS1R* which has been critical to elucidate a new branch of GnRH biology, and for which over 900 manuscripts have been published in the last 13 years. For these instances, complex mechanisms such as oligogenic inheritance, variants pathogenicity and mutational hotspots would be the “missing pieces” to differentiate disease-associated genes with false positives. Interestingly, a recent report showed no enrichment in *KISS1* rare variants in 1,025 GnRH deficient patients, however no categorization for subgroups was used. (i.e., patients with KS, nCHH, hypothalamic amenorrhea and CDGP)¹⁵⁵. Conversely, when analyzing patients with CHH and cerebellar ataxia, another recent study demonstrated a significant enrichment of rare variants in the *RNF216* gene in patients vs. controls⁷¹. Combined, the results of the current study along with results from previous studies indicate the importance of phenotypic clustering to identify novel associated genes^{60,69,85}.

The combination of mutations in both *FGFR1* and *CHD7* occurred most frequently (4 probands). These two genes might play coordinated roles during GnRH neuron development and migration, as *CHD7* regulates the transcription of *Fgf8*, a major ligand for *FGFR1* in GnRH neuron ontogeny⁶⁵. Both *FGFR1* and *CHD7* are expressed in relevant tissues for CHH, such as the olfactory bulb and hypothalamus^{156,157}. A previous report also suggested functional interactions between

these genes, as CHH patients with mutations in *FGFR1* and *CHD7* exhibit overlaps in associated phenotypes (cleft lip/palate, coloboma or ear anomalies)¹⁵⁸.

One notable exception to oligogenic interactions was *ANOS1* – which was inherited in an exclusively monoallelic fashion, due to its X-linked recessive mode of inheritance and high penetrance. Mutations in other genes such as *TAC3*, *KISS1*, *PROK2*, and *PROKR2* were primarily biallelic, and oligogenic interactions were not observed – likely due to their recessive mode of inheritance. Interestingly, the frequency of monogenic inheritance in KS was significantly higher than in nCHH. To date, it is unclear whether this difference is due to distinct genetic architecture or that “missing” partners in an oligogenic inheritance for KS have yet to be discovered.

Putatively pathogenic mutations in CHH genes were found in 17% of control cohorts, which at first glance seems counterintuitive. Importantly, oligogenic inheritance was rarely found in controls (2%), supporting the model of oligogenicity in the pathogenicity of CHH. Additionally, many of the putative heterozygous mutations in controls were found in genes with an autosomal recessive inheritance, and thus would explain the lack of obvious reproductive phenotypes among controls. CHH and controls differ markedly for protein-truncating variants (PTVs) (29% versus 5%, respectively), and the PTVs in controls were less likely to be pathogenic.

The goal of this study was to characterize the genetic overlap of CHH and CDGP, screening the known genes implicated with the two disorders. Despite having full exome data available from a cohort of singleton patients, a parallel analysis to identify new candidate genes involved in the pathogenesis of CDGP was not taken into account.

Future projects will aim to do so, with the involvement of nuclear families and/or siblings to increase the yield of genetic diagnoses in these patients.

In conclusion, the genetic profiles of CHH and CDGP appear to be distinct. This observation may facilitate differential diagnosis of CHH and CDGP in adolescence when a clear and early diagnosis is critical to initiate timely induction of puberty in patients with CHH. A genetic test resulting in 1) more than one CHH gene mutated (oligogenicity), 2) hemizygous *ANOS1* mutations in male patients, or 3) biallelic mutations in genes associated with autosomal recessive inheritance would favor a diagnosis of CHH. This study expands our understanding of the genetic architecture of both CHH and CDGP, and additional comprehensive studies in larger cohorts may enable genetic testing to inform a more precise differential diagnosis in the clinical setting.

Section 3

Biology-driven identification of new CHH genes

Background and Rationale

The complex genetic architecture of CHH including oligogenicity, variable expressivity and incomplete penetrance, constitute important limiting factors in the identification of new CHH loci through WES. In this perspective, biological information could be instrumental to overcome this limitation. Here, I have used a biology-driven strategy based on gene function of well-known CHH genes.

Anosmin-1 is encoded by *ANOS1*, the first gene reported to be implicated in CHH with anosmia (Kallmann syndrome, KS) which plays a critical role in the migration of GnRH neurons from the olfactory placode to the brain during development¹⁵⁹. Anosmin-1 contains a fibronectin type-III (FN3) domain, an adhesion motif found in a wide range of extracellular matrix proteins¹⁶⁰. The majority of *ANOS1* mutations reported in KS patients are located within the FN3 domains suggesting a critical role of these motifs in GnRH biology. Later, mutations in additional FN3-domain encoding genes such as *AXL* and *FLRT3* were identified as causative of CHH^{70,60}. Mutations in FN3-encoding genes were reported in other human diseases associated with altered neuronal development, such as L1 syndrome caused by mutated *L1CAM*¹⁶¹. Other genes encoding for FN3 domains proteins play important roles in GnRH neuron biology in murine studies.

Ephrins also contain FN3 domains and are ligands for receptor tyrosine kinases involved in neuronal migration during early development. In particular, Ephrin A5 (*EPHA5*) exhibits critical functions in the migration of GnRH neurons in mice¹⁶². Another example are the neural cell adhesion molecules (*NCAM1* and *NCAM2*), which contain FN3 domains and directly guide GnRH neurons through the olfactory axons¹⁶³.

In combination, our knowledge of genes encoding FN3 domains suggest that additional genes encoding FN3 domain proteins will be mutated in CHH + anosmia (KS) through their role as guidance molecules during the migration of GnRH neurons from the olfactory placode to their final location in the hypothalamus.

Methods

Cohort description

The cohort consisted of 133 probands (73 KS and 60 nCHH), including 90 males and 43 females, and family members when available.

Bioinformatic filtering of rare variants

A comprehensive list of genes (n=175) encoding for FN3 domains (PF00041) was extracted from the Pfam database (<http://pfam.xfam.org/family/FN3>) and served for the basis of the subsequent filtering processes. Rare (MAF<0.1%) PTVs, inframe InDels, and missense variants predicted to be deleterious in SIFT and/or PolyPhen-2 were cross-referenced with the list of genes encoding for FN3 domains. Genes known to be involved in axon guidance according to the GeneOntology database (GO:0007411) were further prioritized. Additionally, candidate genes were prioritized using OMIM database, according to their implication in human genetic diseases characterized by alterations in central nervous system neurobiological processes or CHH associated phenotypes (i.e., sensorineural hearing loss, cleft palate, dental agenesis, renal agenesis, cerebellar ataxia, skeletal malformations, mirror movements).

Human genetic studies of putative mutations

Mutations in prioritized genes were confirmed by Sanger as previously described⁶⁰. Segregation in available family members was assessed and pedigrees were constructed.

Immunohistochemistry of human brain tissues

In collaboration with Paolo Giacobini, PhD, University of Lille, human fetuses (10 weeks post-amenorrhea n=2) were obtained from voluntarily terminated pregnancies with the parent's written informed consent (Gynecology Department, Jeanne de Flandre Hospital, Lille, France, protocol n°: PFS16-002). The studies on human fetal tissue were approved by the French agency for biomedical research (Agence de la Biomédecine, Saint-Denis la Plaine, France, protocol n°: PFS16-002). Fetuses were fixed by immersion in 4% paraformaldehyde (PFA) at 4°C for 3 days. The tissues were cryoprotected in 30% sucrose/PBS at 4°C overnight, embedded in Tissue-Tek OCT compound (Sakura Finetek, USA), frozen in dry ice and stored at -80°C until sectioning. Frozen samples were cut serially at 20 µm using a Leica CM 3050S cryostat (Leica Biosystems Nussloch GmbH, Germany).

Immunohistochemistry for GnRH was performed as previously reported (Casoni et al., 2016) using a guinea-pig anti-GnRH (EH#1018; 1:10000), produced by Dr. Erik Hrabovszky (Laboratory of Endocrine Neurobiology, Institute of Experimental Medicine of the Hungarian Academy of Sciences, Budapest, Hungary) and previously characterized in post-mortem human hypothalami¹⁶⁴. Antibodies against DCC¹⁶⁵ (Goat IgG, sc-6535; Santa Cruz) and Netrin-1¹⁶⁶ (Monoclonal Rat IgG2A; MAB1109; R&D System) were used at a dilution of 1:500. Samples were rinsed in TBS (TRIS buffer

saline, pH 7.6) and subsequently blocked for 2-hrs at room temperature in blocking solution (TBS+0.3% Triton X-100; Sigma), 0.25% BSA (Bovine Serum Albumin; Sigma) and 10% normal donkey serum (D9663; Sigma). Sections were incubated in a cocktail of primary antibodies (guinea pig anti-GnRH, rat anti-Netrin-1, goat anti-DCC) diluted in blocking solution for 48-hrs at 4 °C. Sections were then rinsed in TBS and incubated in a cocktail of fluorochrome-conjugated secondary antibodies (all raised in donkey; Alexa-Fluor 488-, 568-, 647-conjugated secondary antibodies; Molecular Probes, Invitrogen) diluted at 1:400 in TBS for 2 hrs at room temperature. Sections were then rinsed in TBS, coverslipped with Mowiol, and imaged using an inverted laser scanning Axio observer confocal microscope (LSM 710, Zeiss; Imaging Core Facility of IFR114 of the University of Lille 2, France).

Functional characterization of variants and cell culture

In collaboration with Justine Bouilly, PhD, CHUV, site-directed mutagenesis was used to generate all variants using QuickChange XLII Kit (Stratagene) and confirmed by Sanger sequencing. Primers flanking the mutations were used for subsequent PCR amplifications (5'→3').

CHO and COS7 cells were grown in DMEM-F12 and DMEM-high glucose medium, respectively, and supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine and antibiotics (Gibco, Carlsbad, CA, USA) in humidified air containing 5% CO₂ at 37°C. Transient transfections of cells were carried out using FuGENE 6 reagent (Promega), according to the Manufacturer's protocol.

After transient transfection in CHO cells, protein expression of tagged wild-type and mutated proteins was detected by Western blotting. Briefly, cells were lysed in RIPA buffer and separated with 8% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by electro-transfer onto a nitrocellulose membrane. For immunoblotting, the membranes were probed overnight at 4°C with anti-HA (Roche) and anti-V5 (Invitrogen) antibodies and then incubated with appropriate infrared fluorescent secondary antibodies (Licor).

Total RNAs were extracted from CHO cells after transient transfection using Trizol reagent (Ambion) according to the manufacturer's recommendations and processed for RT-PCR as previously described¹⁶⁷.

Cells were transfected with plasmids encoding the different wild-type and mutant DCCs, Netrin-1, appropriate firefly luciferase reporters, and β -Galactosidase (internal control) in 96-well plates, using the FuGENE 6 transfection reagent (Promega). Luciferase activity was measured using NOVOstar (BMG LABTECH), normalized to β -galactosidase activity and analyzed using Synergy™ Mx (BioTek). All experiments were performed three times in sextuplicates and are expressed as relative light units (RLU).

Briefly, COS7 cells were plated in 24-well dishes and cultured in DMEM-high glucose supplemented with 10% FBS and penicillin/streptomycin. After 24h, cells were transfected with 1 μ g of control vector, pCMV-hDCC-HA WT or mutants using Lipofectamine 2000 reagent (Invitrogen). Culture media was removed 24h after transfection and replaced by Netrin-1-V5 protein in DMEM; 10% FBS for 2h at 37°C. Cells were then washed, fixed 5 min with 4% PFA, blocked using PBS-10% goat serum, and permeabilized with 0.1% triton X-100. Bound Netrin-1-V5 protein was detected

using a monoclonal primary antibody (Invitrogen) in 10% goat serum-PBS and a horseradish peroxidase (HRP)-conjugated goat anti-mouse secondary antibody in 1% goat serum-PBS (Thermo Fisher). HRP activity was revealed using SigmaFast OPD peroxidase substrate (Sigma) and read at 492 nm using a microplate spectrophotometer. The signal is calculated relative to protein concentration. 48h post-transfection, CHO cells expressing WT or mutated DCC were treated with Netrin-1 WT or Mock-CM. DCC/Netrin-1 binding was revealed by ELISA assays and expressed in relative units. Conditioned-media, collected 24h after transfection with plasmids encoding WT, mutated Netrin-1 or empty vector, were used to treat DCC-WT transfected CHO cells. DCC/Netrin-1 binding was revealed by ELISA assays and expressed in relative units. CHO cells were transiently co-transfected with Egr1-luciferase reporter and B-galactosidase control plasmid plus different combinations of plasmid encoding WT Netrin-1 and DCC mutants or *vice versa*. ERK signaling was observed as luciferase activity above baseline, defined as the activity observed with the empty vector alone.

Results were presented as mean \pm SEM of 3 independent experiments each performed in sextuplicate. Statistical significance was evaluated with ANOVA test. $P < 0.05$.

Results

CHH probands harbor rare variants in DCC and its ligand NTN1

Variants in 133 CHH patients were filtered to include 1) genes encoding FN3 domain-proteins involved in axon guidance (GO:0007411), 2) rare variants (MAF<0.1% in ExAC), and 3) PTVs, inframe InDels or missense variants predicted to be deleterious by SIFT and/or PolyPhen-2. Based on this analysis, 18 genes were selected. Four of these genes – including *ANOS1* – are implicated with human neurological disorders (Table 5).

Several lines of evidence pointed to *DCC* as a putative novel gene implicated in CHH: 1) human genetic studies had identified *DCC* mutations in large kindreds exhibiting synkinesia (mirror movements), a phenotype often associated with KS^{72,168}, 2) studies have demonstrated the key roles of the DCC/Netrin-1 complex in axonal growth¹⁶⁹ and 3) *Dcc* knockout mice exhibit impaired GnRH neuron migration^{170,171}.

Table 5. Genes encoding FN3-domain proteins implicated with neurological disorders in OMIM.

Gene	Human Diseases in OMIM	Inheritance
<i>ANOS1</i>	<u>Hypogonadotropic hypogonadism 1 with or without anosmia (Kallmann syndrome 1)</u>	X-linked
<i>CNTN2</i>	Epilepsy, myoclonic, familial adult, 5	Autosomal recessive
<i>DCC</i>	<u>Mirror movements 1</u>	Autosomal dominant
<i>L1CAM</i>	Corpus callosum partial agenesis; MASA syndrome; Hydrocephalus with Hirschprung disease	X-linked recessive

A total of four heterozygous *DCC* missense variants (p.Asn176Ser, p.Pro645Ser, p.Gly649Glu, p.Ser876Tyr) were identified in the cohort. Two variants were not observed in ExAC controls. *DCC* encodes for *Deleted in colorectal carcinoma*, the receptor for Netrin-1 involved in neuronal axon guidance in the developing nervous system. *DCC* is part of the Immunoglobulin superfamily, and has critical roles in

inducing apoptosis when it does not bind its ligand. While only the p.Asn176Ser variant maps to the second Immunoglobulin-like domain of the protein – implicated in protein stability, the remaining 4 variants are located in the FN3 domains. Interestingly, the p.Ser876Tyr variant maps to the fifth FN3 domain, which is implicated in the interaction of DCC with its ligand, Netrin-1¹⁷² (Figure 14).

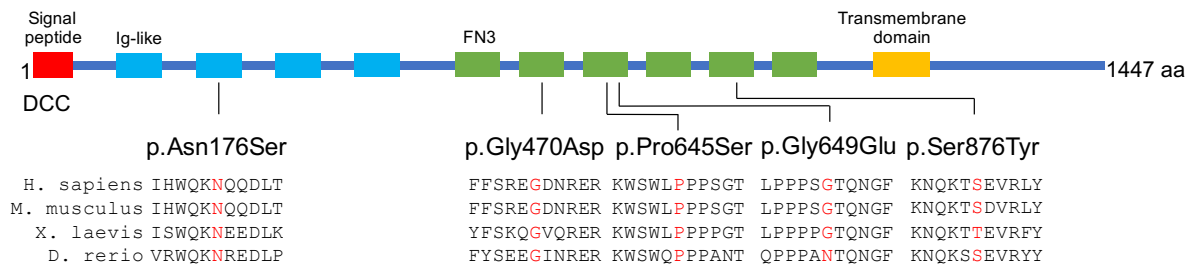


Figure 14. DCC variants located in protein structures with conservation across species.

Given this relationship, the screening was then expanded to *DCC*'s cognate ligand, *NTN1* and one heterozygous missense variant was identified in the NTR-like domain. The *NTN1* variant was absent from both ExAC and CoLaus controls (Figure 15).

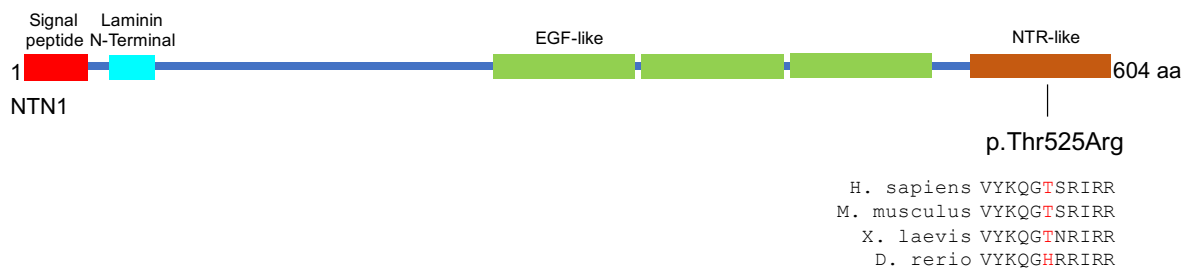


Figure 15. NTN1 variants located in protein structures with conservation across species.

Interestingly, the proband harboring the p.Thr525Arg *NTN1* variant also exhibits a rare variant in *DCC* (p.Gly470Asp; MAF<0.3%). This variant was also identified in his affected KS brother and in an unrelated KS proband. Although our original filtering

strategy for novel gene discovery included a MAF<0.1%, the variant was kept because of digenic inheritance. All rare variants except two (*DCC* p.Gly649Glu and *NTN1* p.Thr525Arg) were predicted to be deleterious by at least one algorithm (Table 6).

Table 6. *DCC* and *NTN1* variants identified in CHH probands.

Gene	Amino acid Change	ExAC MAF	Deleterious predictions
<i>DCC</i>	p.Asn176Ser	0.01%	2/2
	p.Gly470Asp	0.32%	2/2
	p.Pro645Ser	0.02%	2/2
	p.Gly649Glu	private	0/2
	p.Ser876Tyr	private	2/2
<i>NTN1</i>	p.Thr525Arg	private	0/2

ExAC MAF was calculated on ethnicity-matched controls. Deleterious predictions were output by SIFT and PolyPhen-2 algorithms.

***DCC* and *NTN1* variants are loss-of-function**

Western blot analysis demonstrated that protein expression was not affected by the mutations (data not shown). The functional effects of the identified variants were then tested by evaluating the binding properties of DCC to its ligand, Netrin-1¹⁶⁸. Two DCC mutants (p.Asn176Ser and p.Ser876Tyr) displayed impaired binding to wild-type Netrin-1 (Figure 16A), while both Netrin-1 mutants exhibited a significantly reduced binding to DCC WT (Figure 16B). Finally, a transcription reporter assay confirmed that all DCC and Netrin-1 mutants cause significant defects in ERK signaling (Figure 16C-D) — a signaling pathway known for its role in axon guidance and cell migration. Combined, the *in vitro* data demonstrate that the rare variants observed in *DCC* and *NTN1* are loss-of-function mutations.

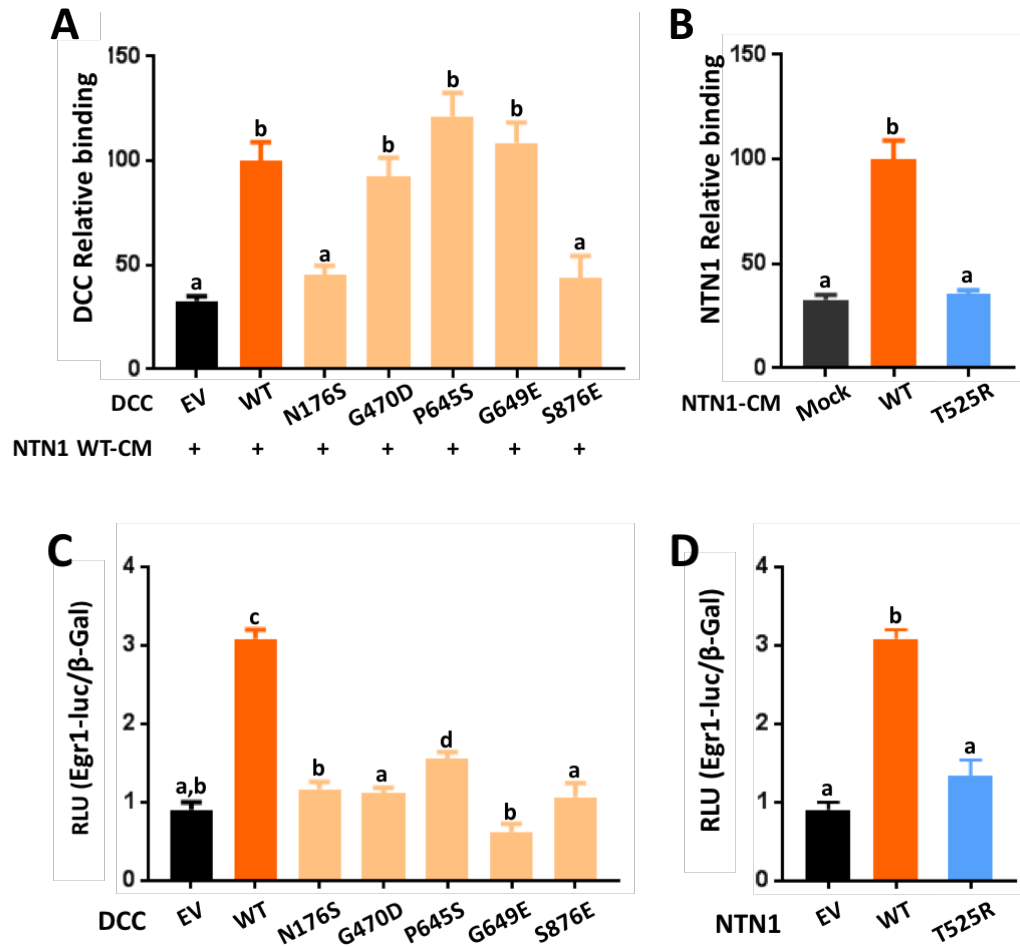


Figure 16. *DCC* and *Netrin-1* variants are loss of function. (A) Analysis of Netrin-1 binding on the mutant and WT *DCC* receptor. (B) Binding analysis of WT and mutant Netrin-1. (C, D) Transcription reporter assay of WT and altered *DCC* or Netrin-1. Different letters indicate significant differences between groups.

Mutations in DCC and NTN1 are associated with severe GnRH deficiency

In total, 3 male and 3 female unrelated CHH probands (4.5%) harbor heterozygous *DCC* and/or *NTN1* mutations (Figure 17, Table 7), and most cases (5/6) exhibit a familial inheritance pattern.

Table 7. Clinical phenotypes of CHH patients with heterozygous *DCC* and *NTN1* mutations.

Family	Subject	Diagnosis	<i>DCC</i> variants	<i>NTN1</i> variants	Sex	Inheritance	Additional phenotypes	CHH genes variants
1	II-1	KS	p.Asn176Ser	-	M	Sporadic	Cryptorchidism, micropenis, hypospadias	<i>PROKR2</i> p.Leu173Arg
2	II-1	KS	p.Gly470Asp	-	M	Sporadic	Micropenis	-

(neonatal)								
3	III-1	KS	p.Pro645Ser	-	F	Familial	-	-
4	II-2	KS	p.Gly649Asp	-	F	Familial	-	<i>CHD7</i> p.Tyr1616Cys <i>SEMA3A</i> p.Arg66Trp
5	II-1	nCHH	p.Ser876Tyr	-	F	Familial	-	-
6	II-1	KS	p.Gly470Asp	p.Thr525Arg	M	Familial	Cryptorchidism	-
	II-2	KS	p.Gly470Asp	p.Thr525Arg	M		Cryptorchidism	

Notably, all probands had absent puberty, and all males had a history of micropenis with or without cryptorchidism, a phenotype associated with severe GnRH deficiency. The majority of probands harboring mutations in *DCC* or *NTN1* (5/6) have KS, based on abnormal olfactory function. This includes a neonatal KS proband (Pedigree 2, Subject II-1) with unilateral right olfactory bulb agenesis on brain MRI.

Previous reports have shown a reversal of CHH in approximately 10% of cases¹⁷³. However, no reversals were observed among probands harboring mutations in *DCC* or *NTN1*. Additional CHH-associated phenotypes in the present cohort were variable, including synkinesia (n=2), bilateral sensorineural hearing loss confirmed by audiogram (n=1) and mild facial asymmetry with unilateral auricular hyperplasia (n=1). Three of the 6 probands with *DCC* or *NTN1* mutations are obese, one has confirmed glucose intolerance, and two presented with reduced bone density (osteopenia/osteoporosis).

Family members were available for genetic analysis in five cases and revealed asymptomatic carriers consistent with reduced penetrance in Pedigrees 1, 2 and 3 (Figure 16). Mutations in other CHH loci in addition to the *DCC/NTN1* mutations (oligogenicity) were identified in Pedigrees 1, 3, and 6 (Figure 17). In Pedigree 1, the proband and the unaffected brother exhibit both a heterozygous mutation in *DCC*

(p.Asn176Ser), and a known founder loss-of-function heterozygous mutation in *PROKR2* (p.Leu173Arg)¹⁴⁸. The parents are asymptomatic carriers of these respective mutations. In Pedigree 3, the proband harbors triallelic mutations in *CHD7* (p.Tyr1616Cys) and *SEMA3A* (p.Arg66Trp) in addition to *DCC* (p.Gly649Glu). *CHD7* and *SEMA3A* mutations are predicted to be deleterious *in silico* (Table 5)^{74,174}. Last, oligogenic combination of *NTN1* (p.Thr525Arg) and *DCC* (p.Gly470Asp) mutations was observed in two monozygotic twins diagnosed with KS (Pedigree 6). DNA from their parents was not available.

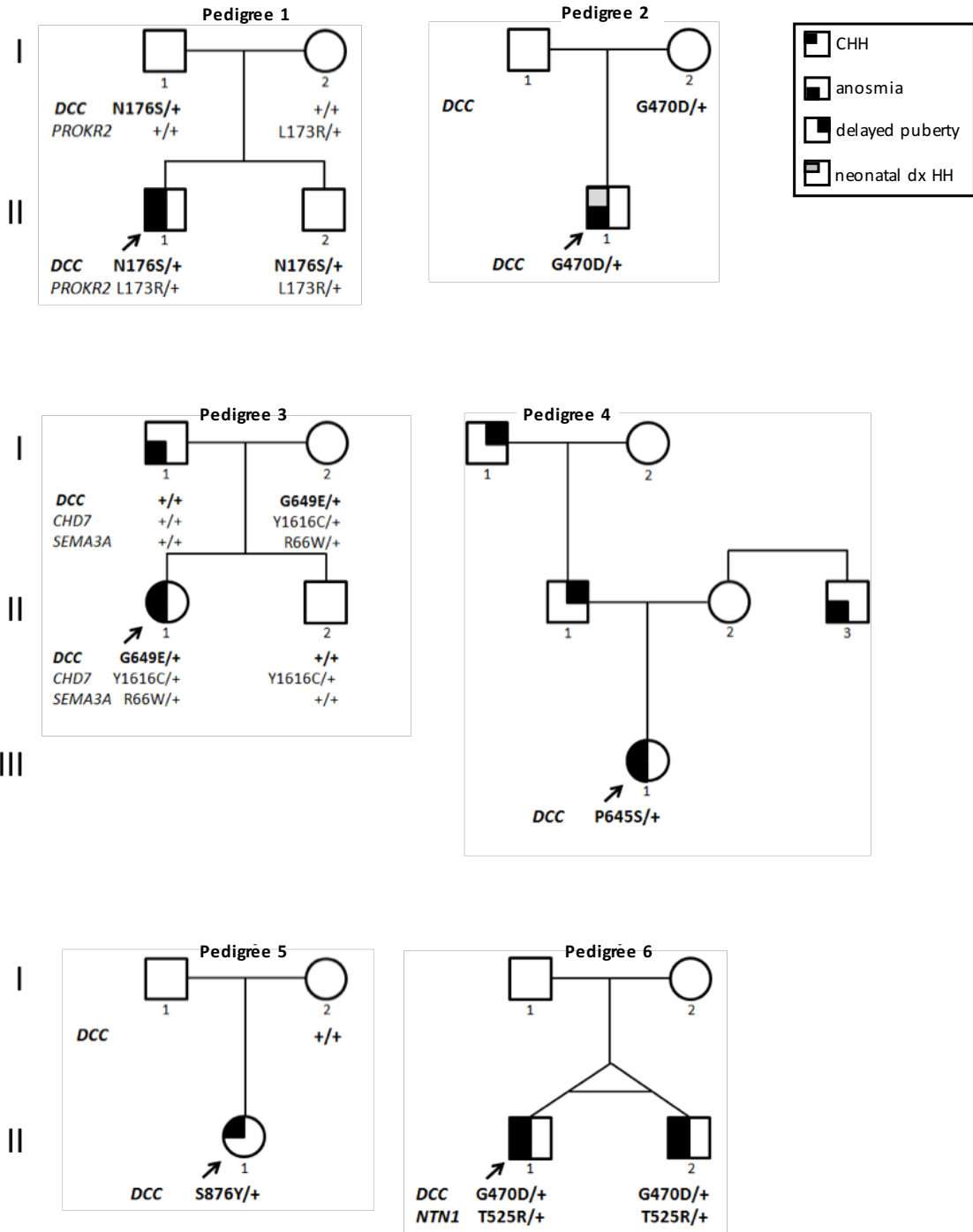


Figure 17. DCC and NTN1 heterozygous variants identified in CHH probands. Pedigree 1, 3, and 6 exhibit oligogenicity.

Netrin-1 and DCC are expressed in human GnRH neurons and the human fetal olfactory system

We then wanted to further support the biological role of *DCC/NTN1* pathway in human GnRH biology. In collaboration with Paolo Giacobini, PhD (University of Lille), we evaluated *DCC* and *NTN1* expression in GnRH neurons and in adjacent tissues during human fetal development (10 gestational week (GW) fetuses) (Figure 18). Simultaneous triple-immunofluorescence experiments on coronal sections of 10 GW fetuses (n=2) revealed that *NTN1* and *DCC* are expressed in the developing vomeronasal organ (VNO) (Figure 18A-C), along the vomeronasal nerve extending from the VNO towards the forebrain (Figure 18B, C), and in GnRH neurons (Figure 18D-I). Moreover, *NTN1* is also expressed in other cell types belonging to the migratory mass (Figure 18F, H), a heterogeneous coalescence of placode-derived and neural crest-derived migratory cells¹⁷⁵. Expression of *NTN1* was detected also in the olfactory epithelium (Figure 18A), consistent with a role of *DCC/Netrin-1* pathway in the migration of GnRH neurons.

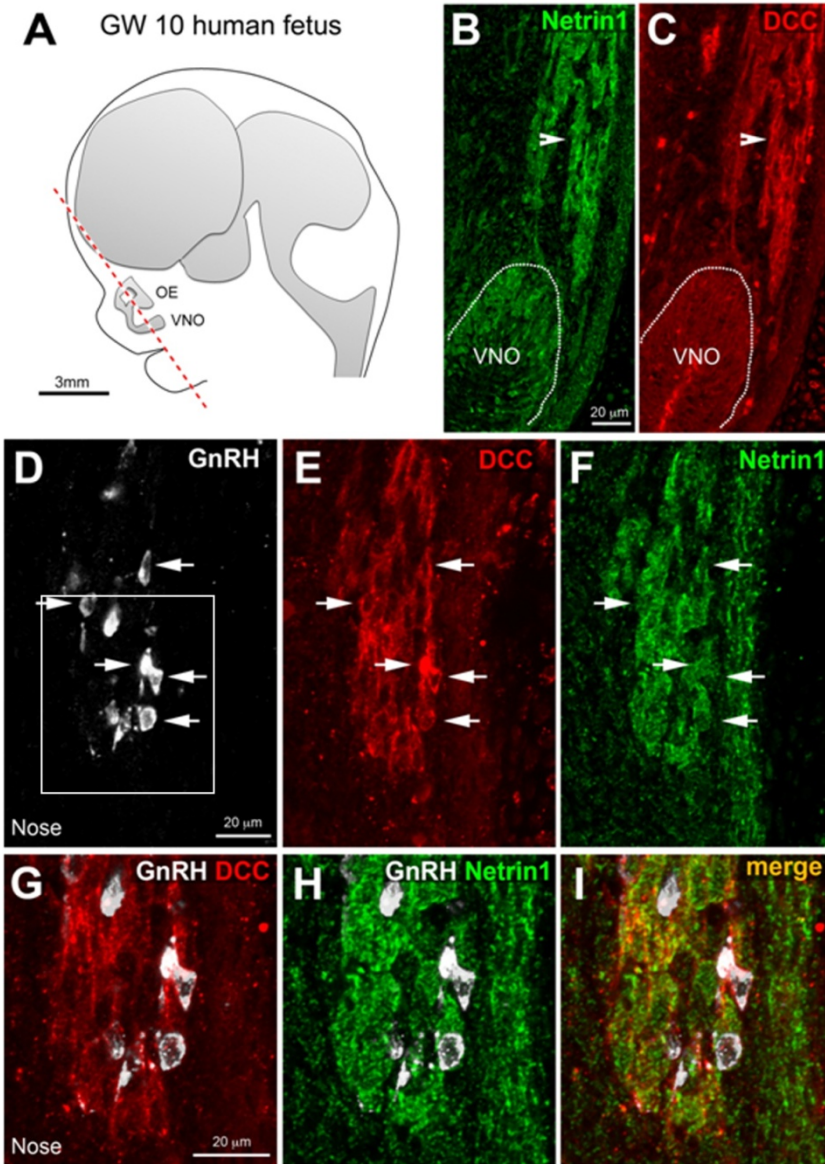


Figure 18. GnRH, DCC and Netrin-1 expression in GW10 human fetus. (A) Schematic representation of a GW10 human fetus head (sagittal view) illustrating the plane of section (red dotted line) used for immunofluorescence. (B) Netrin-1 (green) is expressed in the VNO. Netrin-1 (B) and DCC (red, C) are both expressed in the VNO, VNN and in cells of migratory mass (arrowheads). (D-F) Single-optical confocal planes showing migrating GnRH neurons (white, D) co-expressing DCC (E) and Netrin-1 (F). (G-I) High-power photomicrograph of the region indicated in white box (D), confirming the co-localization of the three antigens. OE: olfactory epithelium. VNO: Vomeronasal organ. Scale bars: 20 μ m.

Discussion

WES coupled with a bioinformatics approach to prioritize FN3-domain containing proteins involved in axonal guidance was used to identify novel potential loci implicated with CHH. This approach was successful with identification of loss-of-function mutations in *DCC* and *NTN1*, which encodes for a ligand of *DCC*. In total, I report loss-of-function heterozygous mutations in *DCC* and *NTN1* in 6 CHH probands with prevalence of 4.5% in our CHH cohort.

Several evidences point out to *NTN1* and *DCC* as promising candidate CHH genes: 1) Netrin-1 and its receptor *DCC* play a critical role in axonal guidance during brain development^{176,177}; 2) mouse genetic studies showed that both *NTN1* and *DCC* are essential for GnRH neuron migration^{170,171,178}; 3) *DCC* mutations are also implicated in human congenital mirror movements (CMM)^{168,179,180}, a CHH-associated phenotype⁷². *DCC* and *NTN1* are expressed along the GnRH neurons migratory route, suggesting an analogous mechanism in human GnRH neuron migration and vomeronasal axon guidance. Consistently, the majority of the CHH probands also present with KS, which is consistent with the role of *DCC* and Netrin-1 in GnRH neuron migration and thus the pathophysiology of KS⁷². Notably, digenic mutations in *NTN1* (p.Thr525Arg) and *DCC* (p.Gly470Asp) were found in two monozygotic twins affected by severe KS associated with cryptorchidism, suggesting a synergistic effect of the two variants. Unfortunately, DNA from the parents was unavailable in this case. Interestingly, two KS probands with *DCC* loss-of-function mutations have CMM, a phenotype thought to be involved with the fibers of the corpus callosum, and previously associated with *DCC*¹⁶⁸. Finally, I showed

that *DCC* and *NTN1* are expressed along the GnRH migratory pathway in human brain development.

Consistent with patterns seen in other known CHH genes, a complex genetic architecture is observed in the families analyzed in this study. Three probands with heterozygous *DCC/NTN1* mutations harbored additional mutations in known CHH genes. In some cases, heterozygous *DCC* mutations were inherited from asymptomatic parents. This makes a difficult task to fully ascertain that heterozygous mutations in *DCC* or *NTN1* may cause alone CHH. The genetic architecture of CHH is complex, and oligogenic inheritance of mutations in the known CHH genes is observed. This can at least partially explain the low penetrance and variable expressivity observed both within and across CHH families⁸². In particular, I observed oligogenic inheritance concomitant with pathogenic variants in known CHH genes. Digenic mutations in *NTN1* and *DCC* were found in two monozygotic twins, affected by severe KS associated with cryptorchidism, suggesting a synergistic effect. Finally, based on the incomplete penetrance and the clinical family histories, an oligogenic inheritance implicating the presence of mutations in unknown CHH loci cannot be ruled out.

A recent report demonstrated that, in addition to the CHH-associated phenotype of CMM, pathogenic mutations in *DCC* cause isolated absence of corpus callosum (ACC) with a complex phenotypic variability¹⁸¹. Similarly, biallelic mutations in *DCC* have been recently found in association with a new genetic syndrome including HGPPS (horizontal gaze palsy, scoliosis), ACC, and midline brainstem cleft¹⁸². The severity (missense vs. PTVs) and inheritance (heterozygous vs. biallelic) patterns of *DCC* mutations could partially explain the phenotypic overlap and variability ranging from CMM, ACC, HGPPS

to CHH, suggesting a significant contribution of oligogenicity and/or environmental factors. Together with these evidences, the results shown in this section illustrate the complexity of *DCC* mutations in the genetics architecture of human disorders of axon guidance and stress the relevance of taking into account the potential physiopathological consequences of these mutations in a genetic counseling perspective.

While homozygous *Dcc* knockout mice harbor GnRH neuron migration defects, previous studies have reported abnormal neuronal axon guidance, ataxia, and absent corpus callosum in heterozygous *Dcc* knockout mice. It is important to note that the reproductive system was not assessed in *Dcc*^{-/-} and *Dcc*^{+/-} mice from these studies, and therefore is a promising avenue for future research. This suggests that heterozygous loss-of-function mutations may be pathogenic *per se*.

For two out of five *DCC* mutations (p.N176S and p.S876Y) and one *NTN1* mutation (p.R362C), located in regions already known to be critical for structural integrity and/or for binding^{172,183-185}, our experiments confirmed that these mutations abolish Netrin-1/DCC interaction. Interestingly, although located in a region with unknown function, T525R *NTN1* mutation also displays a clear loss of binding for DCC, suggesting a novel potential biological function of this NTR-like domain. In correlation with our binding assay results, these mutations showed loss of ERK signaling *in vitro*. Interestingly, we observed a similar ERK signaling defect with DCC mutants along intact binding to Netrin-1, in line with a recent report in which artificial DCC mutants with normal binding to Netrin-1 displayed defective signaling¹⁸³.

In summary, I report loss-of-function mutations in *DCC*, a gene encoding FN3-

domain protein, and its ligand, *Netrin-1* in our CHH cohort. The prevalence of *NTN1/DCC* mutations is similar to another FN3-containing gene, *ANOS1*⁷², suggesting that cumulative prevalence of this family of genes is about 10% in CHH, equivalent to *FGF8/FGFR1*. These latter were reported as the first genes to be associated with both KS and nCHH with a prevalence of 10-12%^{9,51}. The high prevalence of *DCC/NTN1* mutations in CHH, the impaired ERK signaling observed with *DCC/NTN1* mutations and the expression in the developing human GnRH system are consistent with the altered migration of GnRH neurons previously observed in *DCC* and *Netrin-1* knockout mice models^{170,171}, and provide a functional link with the phenotype of our CHH patients. These results further confirm the implication of genes encoding FN3-domain protein in the physiopathology of CHH, KS in particular. Finally, this study highlights the clinical relevance of *Netrin-1/DCC* pathway in CHH, confirms their essential role in signaling in human GnRH biology, and constitutes evidence for the inclusion of *DCC* and *NTN1* in the list of genes implicated with CHH.

Section 4

New CHH genes discovery

with family- and population-based analyses

Background and Rationale

In Section 2, the power of WES to comprehensively identify mutations in known CHH genes was demonstrated, which allowed for an improved characterization of the genetic architecture for this complex disease. In Section 3, a biology-driven bioinformatics analysis was employed to identify new genes implicated in CHH. In the current project, I wanted to use an unbiased genetic analysis to uncover novel genes/pathways and biological processes not previously linked to CHH or GnRH biology.

Historically, one successful strategy to identify likely causative novel genes associated with diseases has been to identify and evaluate *de novo* variants in affected individuals¹⁸⁶. *De novo* variants are compelling candidates to pursue when dealing with rare, sporadic disorders — especially those characterized by low reproductive fitness^{186,187}. In fact, as inherited changes are subjected to evolutionary selection^{188,189}, there is a higher likelihood to identify *de novo* variants with deleterious effects. While the role of *de novo* copy number variations (CNVs) in human diseases has been studied for many years^{190,191}, a comprehensive role of *de novo* SNVs/InDels in severe diseases

has only recently been enabled. The advent of WES allowed for the implication of *de novo* mutations in many rare diseases (e.g. Schinzel-Giedion syndrome¹⁹², Kabuki syndrome¹⁹³ and autism spectrum disorders¹⁹⁴⁻¹⁹⁶). Other strategies often used to identify disease-causing mutations include autosomal recessive analysis — the targeting of homozygous or compound heterozygous variants in affected individuals and inherited from unaffected carriers. X-linked recessive analyses are also advised when evaluating disorders with a predominance of affected male probands, given that these probands would inherit the disorder from their unaffected carrier mothers. This strategy has been successfully employed with disorders such as X-linked intellectual disability (XLID)¹⁹⁷ or CHH (*ANOS1*, *NROB1*)^{57,76}.

One challenge to studying rare disorders, especially those that impact reproductive fitness such as CHH, is the rarity of even small families. Thus, the majority of study subjects are often *singletons*. However, cohorts of *singletons* can be used to identify significant enrichments of individual variants or variants within a particular gene relative to controls, and has been successful in identifying novel genes in rare diseases^{198,199}. In contrast to GWAS studies that utilize SNP arrays to discover genetic associations for common traits, RVAS uses WES or WGS data to identify rare variants with a moderate to high deleterious effect on disease phenotype. Additionally, targeting rare variants in rare diseases allows RVAS to require samples sizes 2-3 orders of magnitude smaller than traditional GWAS²⁰⁰ — meaning that cohorts with <100 probands can generate statistically significant results²⁰¹.

In summary, the three family-based and one population-based strategies are well-suited to the specific disease model of CHH, and therefore are likely to identify novel genes associated with this form of GnRH deficiency.

Methods

Cohort description

The overall cohort consisted of 294 subjects, including 183 probands (103 KS, 77 nCHH, 3 neonatal HH) and 111 family members. Family-based analysis for novel CHH gene discovery was performed on families in which no disease-causing mutations in CHH genes were already identified. Analysis was performed on 23 CHH families (17 trios, 4 quartets and 2 complex families). The majority of probands were KS (n=16), with the remaining probands having nCHH (n=6), or neonatal hypogonadotropic hypogonadism (n=1).

Family-based analysis

De novo analysis was performed using TrioDeNovo²⁰² version 0.04, which uses Phred-scaled genotype likelihoods generated by GATK's HaplotypeCaller. A Bayesian model is then applied to each variant in the proband to estimate the two possible outcomes: 1) the variant follows the Mendelian transmission (i.e. the variant is inherited by the parents), or 2) the variant does not follow Mendelian transmission (i.e. the variant is not present in parents, and therefore is likely to be *de novo*). Variants which met the following criteria were used: 1) Phred-scaled quality of >50, 2) read depth >8x, and 3) MAF <0.01% in ExAC ethnically-matched controls. Variants in genes known to show a high number of false positives due to their presence in highly polymorphic genes or copy number regions²⁰³, and variants in segmentally duplicated regions^{204,205} (identified using the UCSC Human Genome Browser "Segmental Dups"

track, <https://genome.ucsc.edu/>) were excluded from analysis. Paternity was confirmed using all exome variants to calculate relatedness²⁰⁶ using the *relatedness2* function of VCFtools²⁰⁷.

The raw sequence alignment files (BAM files) were visually evaluated using the Integrative Genomics Viewer (IGV)²⁰⁸ for probands and their parents. *De novo* variants were considered as passing if the heterozygous variant was clearly visible in the proband and absent in both parents. All passing *de novo* variants were verified by Sanger sequencing using previously described methods⁶⁰.

Autosomal recessive and X-linked recessive analyses were performed using rare variants (MAF<0.1%) with genotype Phred-scaled quality scores >50. Variants in polymorphic genes and segmental duplications were excluded. The analysis was conducted with the GEMINI (GEnome MINing) version 0.11 software. Specifically, this software identifies variants which are heterozygous in unaffected parents, and either homozygous (using the *autosomal_recessive* function) or compound heterozygous (using the *comp_hets* function) in affected probands. Last, the *X_linked_recessive* function filters all the variants showing “homozygous” genotypes in male (i.e., hemizygous) affected individuals, inherited from the unaffected mother.

Array-CGH in CHH trios

Array-CGH analysis was performed in patient-parents trios on the Affymetrix CytoScan HD array platform, with an average resolution of ~20 kb according to the Human Genome build hg19. Cytoscan HD array includes 2.6 million markers, including 750,000 genotypeable SNPs and 1.9 million non-polymorphic probes. Analysis was

carried out following the manufacturer's protocols, and array data were visualized on the Chromosome Analysis Suite (ChAS) version 3.1.

Segregation analysis

Given the incomplete penetrance, variable expressivity and oligogenicity observed in CHH, unaffected and partially-affected (i.e. CDGP, partial CHH, anosmia only, etc) family members were not used for segregation analysis. Variants were evaluated for segregation with the phenotype using 17 pairs of affected relatives. This included 12 affected sibling pairs, and 5 affected parent-child pairs. Segregation was noted to be discordant only when the identical variant was not shared by the affected pairs.

Variants and genes prioritization

Clinical significance of the variants identified in family-based analyses was annotated according the American College of Medical Genetics and Genomics and the Association of Medical Pathology (ACMG/AMP) guidelines²⁰⁹ using InterVar²¹⁰. Variants deemed to be "Pathogenic/Likely Pathogenic" were retained for gene prioritization steps. Genes were prioritized if they 1) are implicated in CHH overlapping phenotypes in human (OMIM, www.omim.org), 2) underlie CHH overlapping phenotypes in mouse models (MGI, www.informatics.jax.org), 3) have additional CHH patients harboring rare variants, 4) have absent/low frequency of homozygous/compound and heterozygous/hemizygous variants in both CoLaus and ExAC controls, 5) have a high constraint for missense variants in ExAC individuals (z score >1.96)¹¹², and/or 6) show

differential expression in our RNA-seq dataset of migratory and post-migratory GnRH neurons in mouse (see below).

RNA-seq in GnRH-positive and GnRH-negative neurons

Noses and basal forebrains from *Gnrh::Gfp* mouse embryos (E13.5) were microdissected and enzymatically dissociated using a Papain Dissociation System (Worthington) to obtain single-cell suspensions. Fluorescence-Activated Cell Sorting (FACS) was performed based on measurements of GFP fluorescence (excitation: 488 nm; detection: GFP bandpass 526/21 nm, autofluorescence bandpass 664/20 nm) by comparing cell suspensions from *Gnrh::Gfp* and wild-type animals²¹¹. For each animal, 400 to 800 GFP-positive or 1000 GFP-negative cells were sorted directly.

50 ng of total RNA from each sample were reverse transcribed (Maxima H Minus RT) with individual oligo-dT primers, featuring a 6 nt long multiplexing barcode and template switch oligo (Microsynth). The sequencing library was prepared by tagmentation of full length cDNA with an in-house Tn5 transposase at 55°C for 9 minutes and purified (DNA Clean and Concentration kit) prior to paired-end sequencing using the NextSeq 500 (Illumina).

Reads from barcoded mRNA-seq experiments have two barcodes corresponding to the two levels of multiplexing. The first one is common to standard protocols and is used to separate the libraries. The second is specific to the barcoded mRNA-seq protocol and is used to separate the multiplexed samples from the bulk data. Raw reads were mapped to the mouse genome (GRCm38) using STAR (with default parameters). Samples that did not aggregate enough reads were excluded from further analysis

(<1M). The count table was then obtained using HTseq. Genes with a count per million (cpm) greater than 1 for more than 20 samples were retained. Raw counts were then normalized using the *voom* package in R²¹².

Genome-wide gene-collapsed RVAS

In order to standard our datasets and avoid introducing bias into our results due to different probe sets used by the different cohorts, only variants identified by BEDTools v2.17.0²¹³ as present in overlapping regions covered by Agilent V2 (used in cohort 1 [see Figure 5] and CoLaus) and V5 (used in cohorts 2-5) were used.

A first filtering process removed all variants without a “PASS” filter after HaplotypeCaller genotyping, in order to reduce the signal due to false positives. Two different variants sets were created taking into account ExAC non-Finnish European and CoLaus controls: 1) variants having MAF<1%, and 2) variants with MAF<0.1%. Next, variants were filtered to include only PTVs, inframe InDels and missense variants predicted to be pathogenic by SIFT and/or PolyPhen-2. PLINK/SEQ was used to perform the gene-based association test, using the default “BURDEN” test which calculates the excess of rare alleles in cases compared to controls in a contingency table similar to a χ^2 test¹⁴⁷. To reduce computing time, the genome-wide analysis was performed using a two-step approach, using 1) adaptive permutations, that stop the test if the ultimate empirical p-value shows not to be significant, and 2) 10 million permutations, in order to reach a minimum p-value of 1×10^{-7} , sufficient to identify genome-wide significance in gene-based RVAS, i.e. $p=2.6 \times 10^{-6}$. Gene-based associations were calculated for CHH European probands vs. CoLaus controls, as well as for the KS

and nCHH subgroups separately vs. CoLaus controls. Significant associations were then validated in a second, larger set of controls, using 33,370 non-Finnish European individuals from ExAC using a Fisher's exact test. Genome-wide associations in cases vs. controls were plotted using the "qqman" package in R²¹⁴.

Results

De novo analysis

Using TrioDeNovo in 14 trios with asymptomatic parents, 17 variants in 9 families were identified after filtering as described in Methods. Visual inspection of filtered variants identified 9 probable *de novo* variants (53%), while the remainder were either inherited from one of the two parents, or were false positives due to reads misalignments (Figure S6). All variants were confirmed using Sanger sequencing were confirmed to be *de novo*. Paternity was confirmed in all trios. Array-CGH analysis did not identify any *de novo* CNVs in the trios (Table S6).

Autosomal recessive analysis

Autosomal recessive analysis identified rare (MAF<0.1%) homozygous and/or compound heterozygous variants in 23 families. Of note, one or both parents had a history of delayed puberty in seven families (30% of analyzed pedigrees). In total, 15 compound heterozygous and 6 homozygous variants were identified in 9 families. aCGH did not reveal any CNVs segregating in an autosomal recessive mode of inheritance (Table S6).

X-linked recessive analysis

X-linked recessive analysis in 18 families with affected male probands identified 29 rare hemizygous variants in 11 male probands. aCGH did not detect any CNVs in probands' X chromosomes inherited from their asymptomatic mothers (Table S6).

In summary, a set of 59 rare variants resulting from the family-based analyses were considered for prioritization steps, in order to identify the most promising novel candidate CHH genes (Table S7).

Variants filtering and prioritization of best candidates

Variants were annotated using InterVar which is guided by the ACMG/AMP recommendations to define each variant's clinical significance. Prior to scoring in InterVar, variants were manually annotated to include whether they were *de novo* or inherited in an autosomal recessive or X-linked fashion as this information plays a vital role in the classification of pathogenicity. In total, 11 of the 59 variants (18%) were classified as "Likely pathogenic" (Table 8).

Table 8. Candidate genes with likely pathogenic variants identified through family-based analysis.

Gene	Inher.	Variant	Segreg. in aff. pairs	ExAC MAF	Other individuals with rare variants		Rare homozygous/compound heterozygous variants			ExAC miss. const.	Human associated diseases	Mouse phenotypes	Diff. exp. in GnRH neurons
					CHH	CoLaus	CHH	CoLaus	ExAC				
<i>ASAP3</i>	DN	p.Thr97Ala	-	private	12	3	1	0	0	0.87	-	-	NS
<i>BARHL1</i>	DN	p.Arg182Leu	-	private	0	0	0	0	1	3.22	-	deafness	ND
<i>SMC3</i>	DN	p.Cys549Tyr	-	private	0	0	0	0	0	6.25	<u>Cornelia de Lange syndrome</u>	abnormal craniofacial morphology	NS
<i>ERCC4</i>	DN	p.Arg740Cys	-	0.004%	6	6	0	0	1	-1.06	Fanconi anemia, complementation group Q; Xeroderma pigmentosum, type F/Cockayne syndrome	decreased body weight; abnormal liver morphology	NS
<i>RCAN1</i>	DN	p.Glu135Ala	-	private	0	1	0	0	0	1.12	-	<u>reduced fertility</u>	NS
<i>MGAT1</i>	DN	p.Arg129Trp	-	private	0	0	0	0	0	3.55	-	abnormal oogenesis; <u>decreased litter size;</u> <u>reduced female fertility</u>	NS
<i>POLR3B</i>	AR	p.Phe400Se/ p.Val523Glu	-	private/ 0.05%	0	0	0	0	0	3.43	Leukodystrophy, hypomyelinating, 8, with or without <u>oligodontia</u> and/or <u>hypogonadotropic hypogonadism</u>	-	NS
<i>SLIT2</i>	AR	p.Asp1074His	-	0.06%	2	2	0	0	0	1.71	-	abnormal Purkinje cell dendrite morphology; <u>abnormal axon guidance;</u> <u>abnormal olfactory tract morphology</u>	0.002
<i>GUF1</i>	AR	p.Ala353Val/ p.Ile478Ser	2/2	0.01%/ 0.00%	1	5	0	0	2	-1.1	Epileptic encephalopathy, early infantile, 40	-	NS
<i>C11orf35</i>	AR	p.Glu463Val	-	private	1	1	0	0	0	-1.13	-	-	NS

Prioritized variants and gene information. DN: de novo; AR: Autosomal recessive. Segregation in 17 affected relatives; 2/2: segregating with the disease in two different pairs with a variant in the gene. ExAC MAF was calculated in all ethnicities. Other variants in CHH cohort and CoLaus/ExAC controls were filtered with the same criteria described in Methods. Compound heterozygous in ExAC could not be assessed. ExAC missense constraint evaluated the number of observed vs. expected missense variants in each gene, generating a Z score of its constraint of harboring missense variants. Z scores >1.96 are considered nominally significant. Human associated diseases information is extracted from OMIM. Mouse phenotypes information is extracted from MGI. Differential expression in GnRH neurons shows p-values from our RNA-seq analysis (see Methods and Table S8). ND: not detected; NS: not significant.

These 11 genes containing likely pathogenic variants were then prioritized using 1) the presence of CHH-associated phenotypes in human disease, 2) the presence of CHH-associated phenotypes in knockout mice, or 3) differential expression demonstrated in migratory vs. post-migratory GnRH neurons in our RNA-seq experiment. This resulted in the identification of 4 high priority candidate genes — two from the *de novo* analysis (*MGAT1* p.Arg129Trp and *SMC3* p.Cys549Tyr) and two from the recessive analysis (*SLIT2* p.Asp1074His/p.Asp1074His and *POLR3B* p.Phe400Ser/p.Val523Glu) (Figure 19).

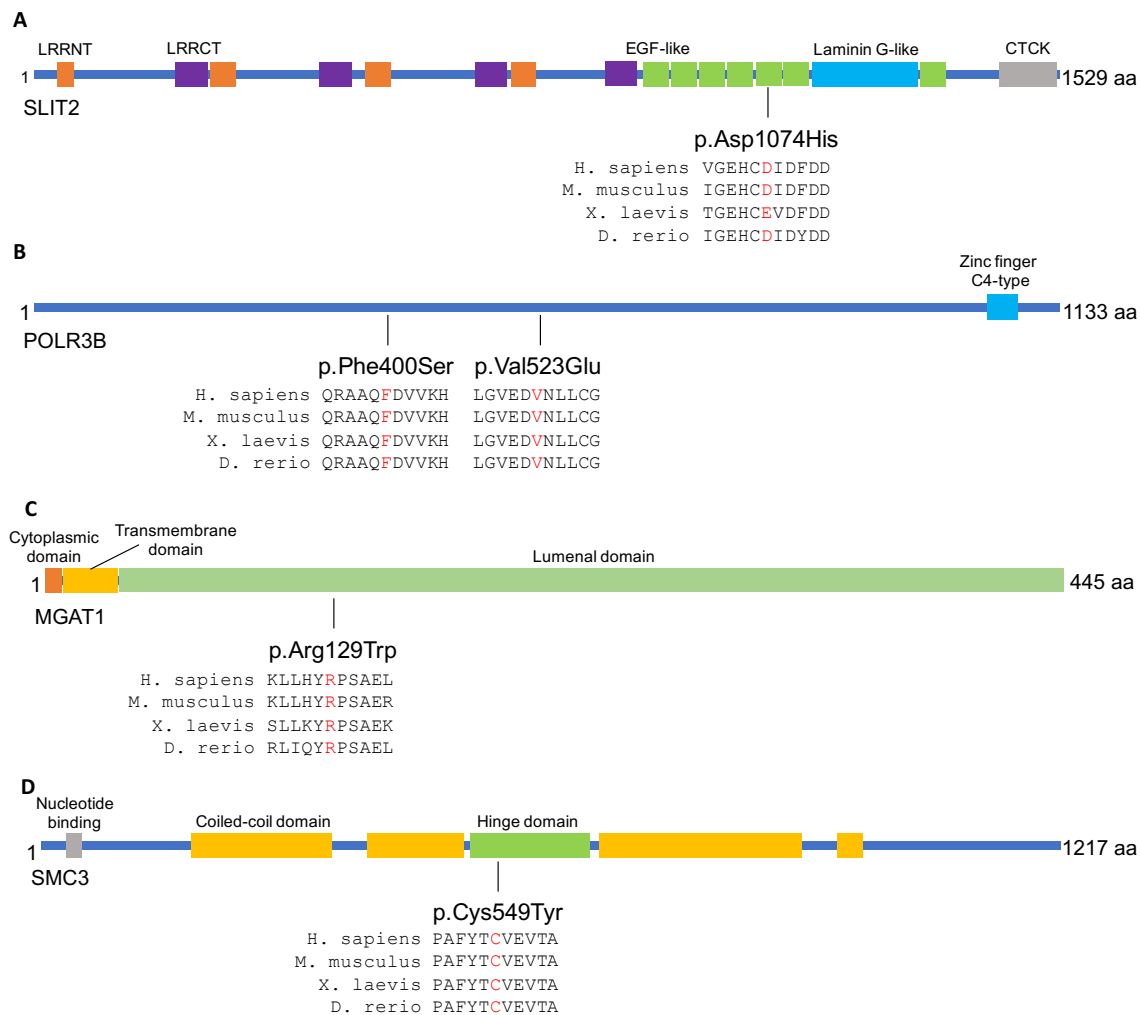


Figure 19. Schematic representing proteins encoded by candidate genes with localization of variants identified in CHH patients.

Characterization of prioritized genes

SLIT2

The homozygous *SLIT2* variant (p.Asp1074His) (Figure 19A) was detected in a KS patient with unilateral cryptorchidism and microphallus (Figure 21, Pedigree A). Additionally, the patient has been diagnosed with Dandy-Walker malformation (macrocephalia, mental retardation, cerebellar ataxia, and agenesis of corpus callosum), as well as hypertelorism, high myopia, high-arched palate, and metabolic defects (insulin resistance and high body fat percentage). Two additional rare heterozygous missense variants (p.Arg225His, p.Thr594Met) were identified in a nCHH and a KS singleton.

SLIT2 encodes for Slit homolog 2 protein, a member of a family of secreted glycoproteins involved in neuronal axon guidance. In mice, Slit2 is highly expressed along the migratory route of GnRH neurons (nasal compartment, forebrain and hypothalamus), and regulates their migration upon binding to Robo3²¹⁵ (Figure 20).

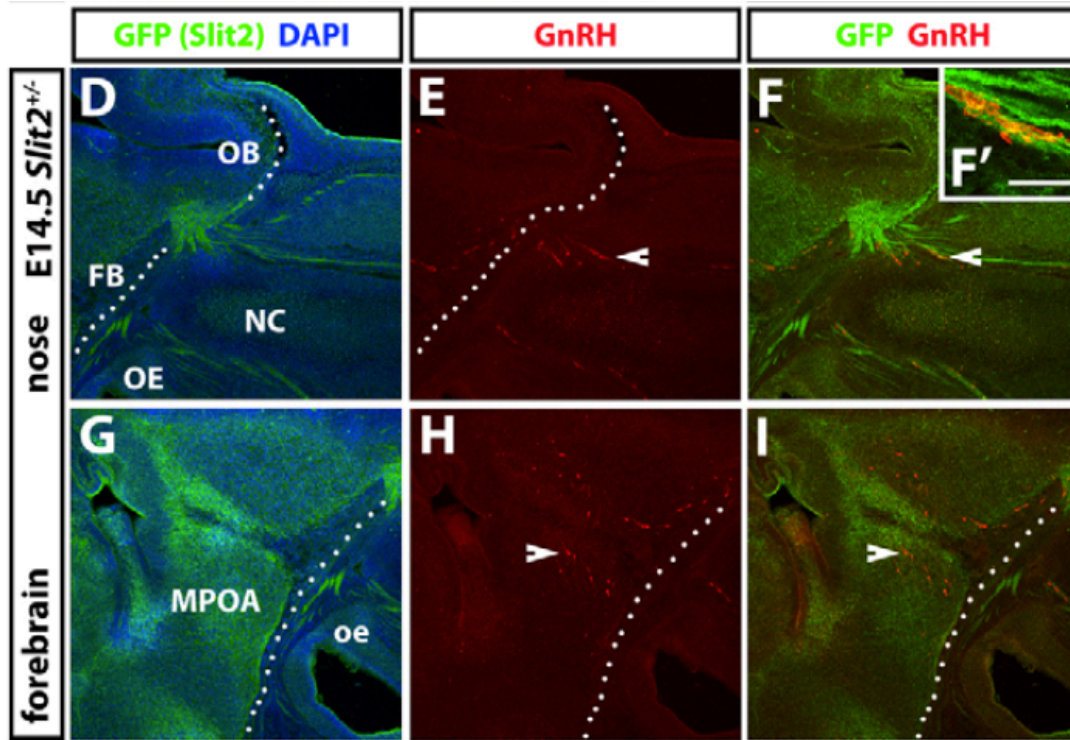


Figure 20. Slit2 and GnRH expression during GnRH neuron migration at E14.5. (From Cariboni et al., 2012)

Additionally, *Slit2*^{-/-} mice showed early GnRH migratory defects, as most GnRH cells fail to reach the hypothalamus and instead accumulate in the nasal compartment²¹⁵. Confirming this observation, *Slit2* was the only gene among the list of filtered genes to show significant differential expression in migratory vs. post-migratory GnRH neurons in our RNA-seq experiment (Table S8).

POLR3B

Compound heterozygous variants in *POLR3B* (p.Phe400Ser/p.Val523Glu) were identified in two nCHH siblings (Figure 21, Pedigree B). *POLR3B* encodes for the DNA-directed RNA polymerase III subunit B involved in the synthesis of small RNAs. No additional rare variants were identified in other CHH probands. Mutations in *POLR3B*

have been implicated in a complex syndromic disorder which is characterized by childhood-onset hypomyelinating leukodystrophies with prominent cerebellar dysfunction, oligodontia, and hypogonadotropic hypogonadism^{216, 217}. The identical combination of compound heterozygous *POLR3B* variants found in our siblings were identified and published in two siblings with CHH and no additional phenotypes²¹⁸. Post-analytical follow-up of our two siblings confirmed this was the identical family published by Richards *et al.*

MGAT1

The *de novo* *MGAT1* variant (p.Arg129Trp) (Figure 19C) was identified in a male KS proband with unilateral cryptorchidism (Figure 21, Pedigree C). The *MGAT1* gene encodes for mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase, a glycosyltransferase involved in the conversion of high-mannose to complex N-glycans. Specifically, *MGAT1* is directly involved in the modification of gonadotropin (FSH, LH, hCG) glycans²¹⁹. Interestingly, the KS patient harboring the *MGAT1* *de novo* missense variant displays resistance to hCG/FSH treatment to initiate puberty. *Mgat1* knock-out mice showed developmental retardation at E10.5 with lack of neural tube formation and neural vascularization, and die at E12.5²²⁰. Oocyte-specific deletion of *Mgat1* in female mice results in infertility and a reduction of ovulated eggs²²¹. Brain-specific *Mgat1* knock-out mice showed neurological defects and live no longer than 18 weeks. Increased neuronal apoptosis has been observed in these mice, and reduced levels of voltage-gated channels proteins were detected (KCNA1, SCNA1, and SCN1B)²²². Mutations in *MGAT1* have not been

implicated in human diseases, and no additional CHH probands harbor rare variants in *MGAT1*. High expression levels in human hypothalamus and pituitary has been observed, and RNA-seq in mouse GnRH neurons detected *Mgat1* expression (Table S8).

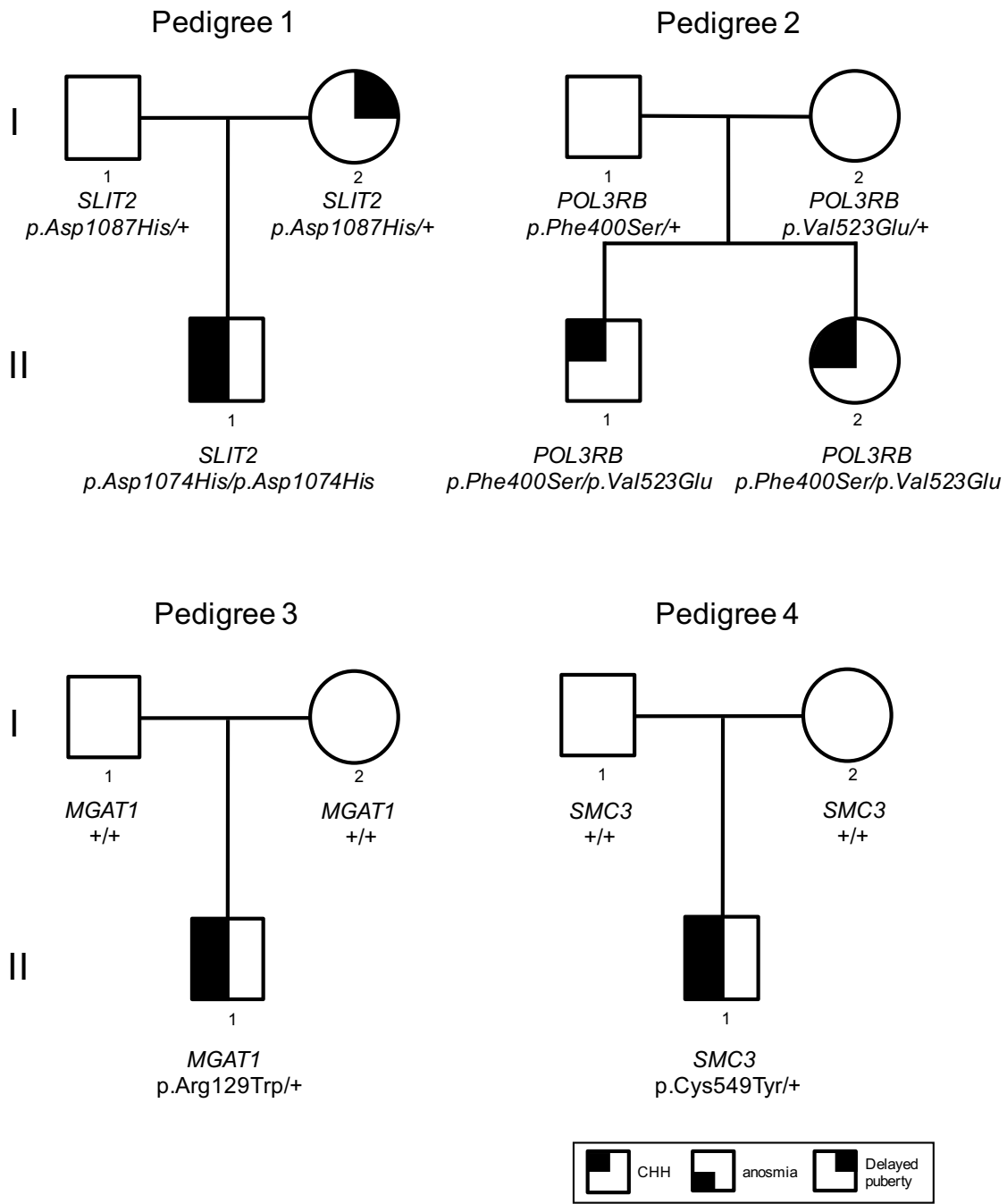


Figure 21. Pedigrees prioritized from family-based analyses.

SMC3

The *SMC3 de novo* missense mutation (p.Cys549Tyr) (Figure 19D) was identified in a KS patient (Figure 21, Pedigree D), and is not present in any of the control databases. *SMC3* encodes for the structural maintenance of chromosomes protein 3, a key component of the cohesin complex. Pathogenic mutations in proteins involved in cohesin function underlie human diseases described as “cohesinopathies”. The two main cohesinopathies are Cornelia de Lange syndrome (CdLS) and Roberts syndrome (RS). CdLS is a rare (1:50,000) autosomal dominant disease mainly characterized by mental and growth retardation, facial dysmorphism, and upper limb abnormalities. Mutations in *SMC3* are found in 2% of individuals with a non-syndromic, milder form of CdLS. Interestingly, 19% of *SMC3*-mutated patients also had symptoms suggestive of hypogonadism such as cryptorchidism in males and late menarche with small breast development in females. CHH-associated phenotypes of hearing loss and clinodactyly of 5th finger were also observed in CdLS patients²²³.

SMC3 expression

SMC3 is expressed in all human tissues according to the GTEx database, as well as in mouse pre-migratory and post-migratory GnRH neurons from our RNA-seq data. *Smc3* expression patterns in the Allen Brain Atlas and GenePaint databases demonstrate it is expressed in the neuroepithelium and olfactory epithelium of mouse embryos at E14.5. The highest expression levels in the adult brain are detected in the olfactory bulb, the rostral migratory stream, sub-ventricular zone, hippocampus and

cerebellum. However, focal expression in specific hypothalamic nuclei, (supra-chiasmatic, paraventricular, and arcuate) involved in energy balance, circadian rhythm and fertility (Figure 22) is also observed.

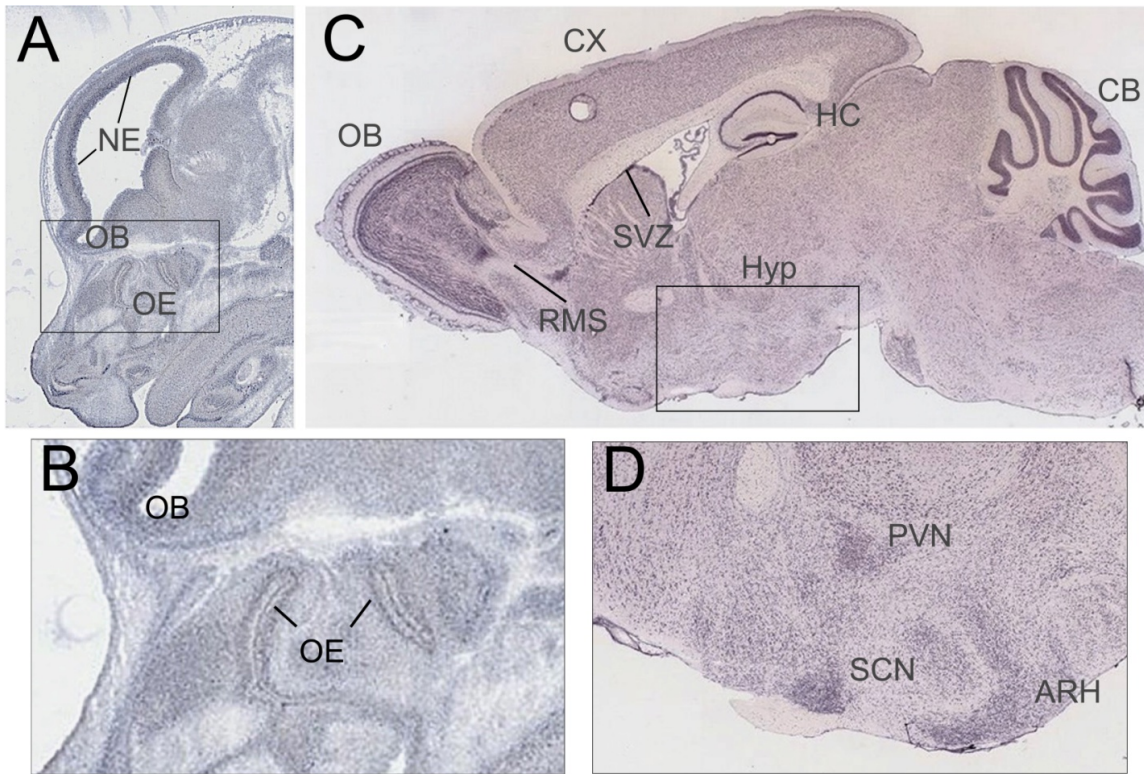


Figure 23. *Smc3* expression in mouse embryo and adult brain.

(A) *Smc3* *In situ* hybridization in E14.5 mouse embryo head sagittal section. (B) Higher magnification of the nasal forebrain junction (black box in A). (C) *Smc3* *In situ* hybridization in adult mouse brain sagittal section. (D) Higher magnification of the hypothalamic region (black box in C). NE, neuroepithelium; OE, olfactory epithelium; OB, olfactory bulb; CX, cortex; RMS, rostral migratory stream; SVZ, sub-ventricular zone; HC, hippocampus; CB, cerebellum; Hyp, hypothalamus; PVN, paraventricular nucleus; SCH, supra-chiasmatic nucleus; ARH, arcuate nucleus.

To better define the *Smc3* expression patterns during development in key tissues for GnRH biology, I performed semi-quantitative RT-PCR in mouse tissues and cell lines. First, *Smc3* expression levels in embryonic development and postnatal stages were compared. *Smc3* levels are significantly higher in E14.5 embryos when compared to

postnatal and adult mice, supporting a crucial role during development consistent with its role in cell replication. No differences in the expression between embryonic nose and brain were observed, however a significant decrease of *Smc3* levels in the preoptic area of the hypothalamus from birth (P0) to adulthood was detected. The broad expression of *Smc3* the brain was confirmed, but notably the olfactory bulb (an area where neurogenesis also occurs during adulthood) showed a significantly higher abundance of *Smc3* (Figure 23).

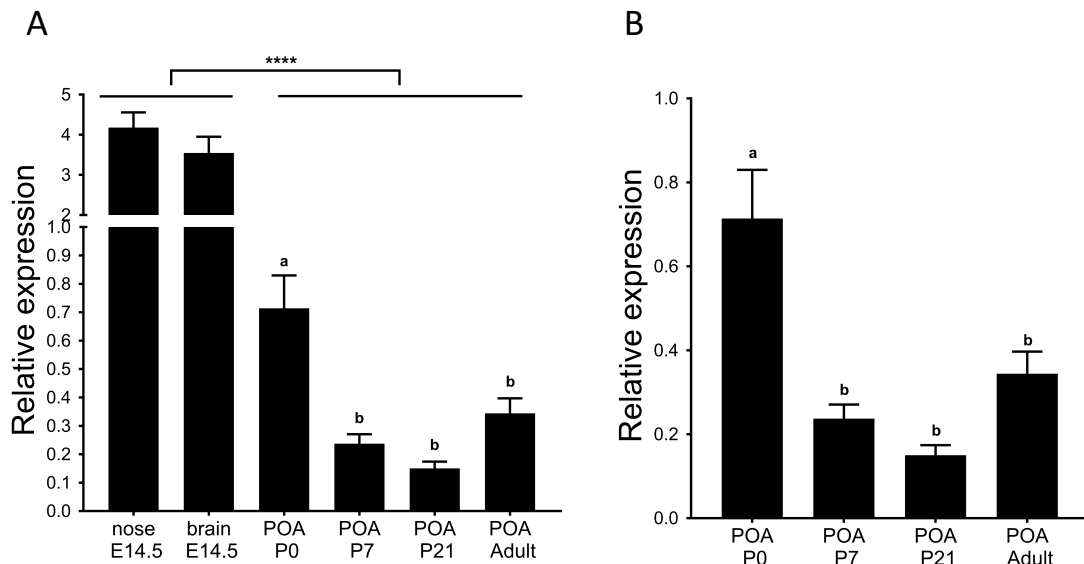


Figure 23. *Smc3* quantitative expression in mouse. (A) Expression levels in nose and brain at embryonic stage E14.5 and in hypothalamic preoptic areas from birth (P0) to adulthood (P56); (B) Expression levels at adult stage in different brain areas (preoptic area of the hypothalamus, olfactory bulb, medial basal hypothalamus, median eminence of the hypothalamus, cortex). Statistical significance was evaluated with ANOVA test. **** = $p < 0.0001$; a vs. b = $p < 0.05$

p.Cys549Tyr mutation impairs SMC3 function

The *de novo* p.Cys549Tyr variant lies within the hinge domain of SMC3, which interacts with its SMC1 counterpart to assemble the cohesin ring. In somatic cells, the SMC3 and SMC1A hinge domains bind to forming an inverted V-shaped heterodimer.

RAD21 interacts with the ATPase heads of both SMC3 and SMC1A. Stromal antigen 1 (SA1) or 2 (SA2) then bind to RAD21, forming the ring-shaped structure of cohesin (Figure 24). The SMC1-SMC3-RAD21 complex functions as a ring to surround DNA fibers, while SA1-2 proteins are crucial for cohesin loading on chromatin²²⁴.

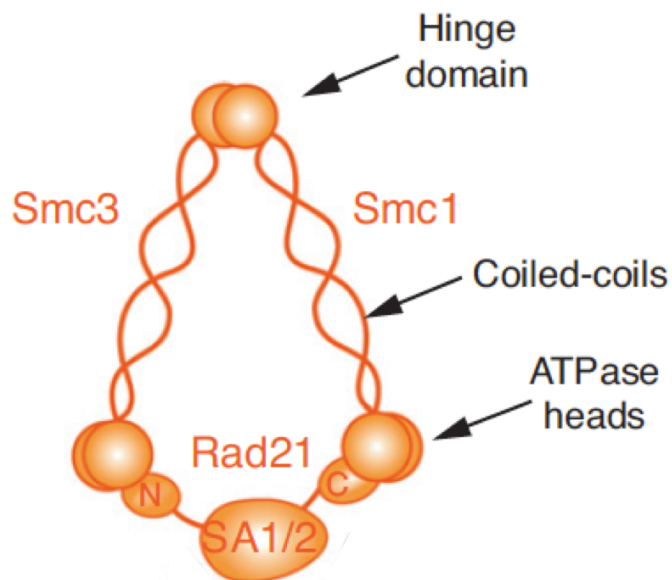


Figure 24. Cohesin complex with SMC3, SMC1, RAD21 and SA1/2 proteins.

In collaboration with Jérôme Fagart (INSERM U1185, Kremlin-Bicêtre, France), the crystal structures of SMC3/SMC1B hinge domains heterodimer were used to predict the effect of the p.Cys549Tyr variant. The cysteine is located within a hydrophobic environment, surrounded by methionine, leucine and phenylalanine (Figure 25). The aromatic ring of the substituted tyrosine residue would disrupt the hydrophobic site resulting in the potential mis-folding/destabilization of the SMC3/SMC1B hinge binding due to steric hindrance and/or the formation of new interactions.

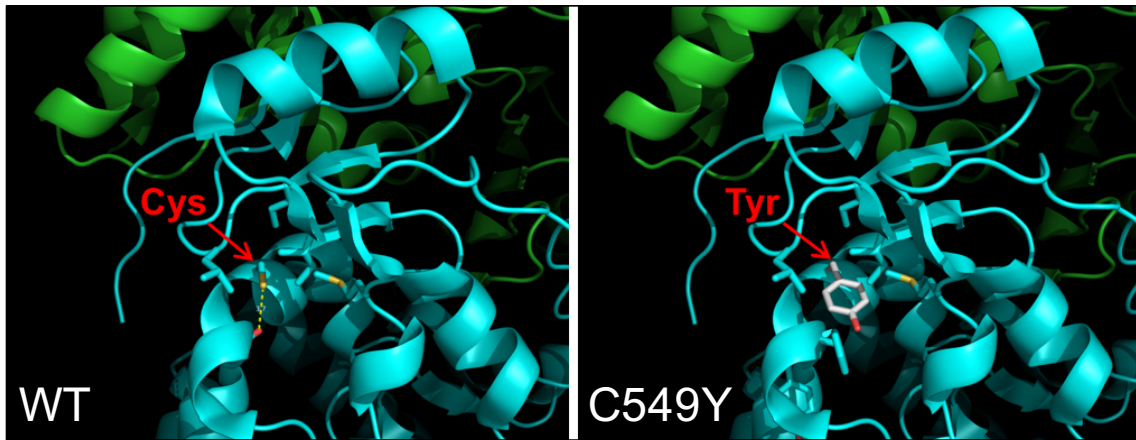


Figure 25. Predicted structural conformations of WT and mutated SMC3 hinge domains.

To test these hypotheses, a yeast two-hybrid assay was performed in collaboration with Justine Bouilly, PhD (CHUV). Wild type and mutated SMC3 hinge domains were transfected with wild type SMC1b hinge counterparts. SMC3 hinge domains were cloned into pBind plasmids, encoding for the binding domain of GAL4 promoter, while SMC1b hinge was cloned into pAct plasmids, encoding for the activator domain of GAL4 promoter. The binding of the two hinges activate the transcription of GAL4 resulting in the production of luciferase. A significant increase of luciferase in mutant SMC3 interacting with SMC1b was observed compared to the interaction of wild type SMC3 with SMC1b (Figure 26). This result shows that the SMC3/SMC1B complex is more stable with the introduction of the tyrosine and indicates of a gain-of-function mutation.

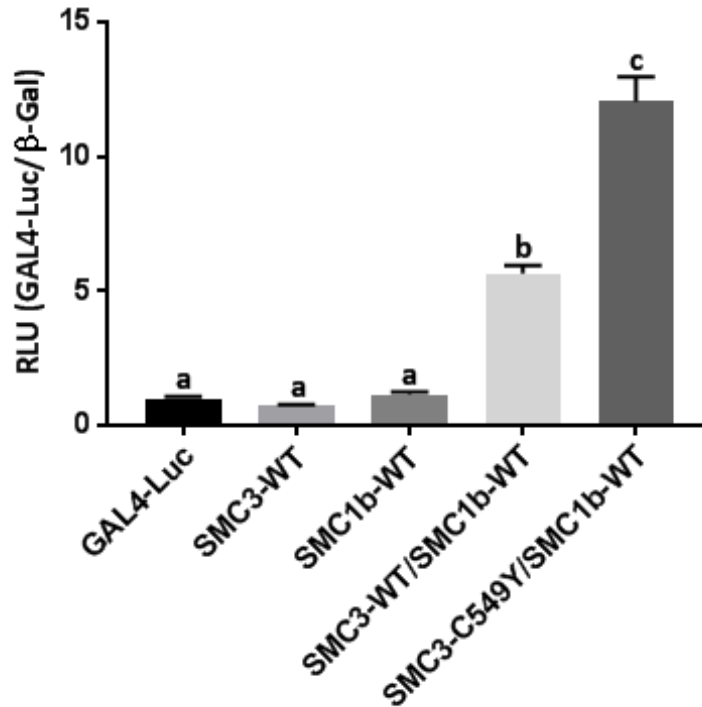


Figure 26. Double-hybrid assay experiments on WT and mutated SMC3 hinge vs. SMC1 hinge.

Potential genetic and phenotypic overlap of CHH and CdLS

The patient harboring a *de novo* missense variant in *SMC3* was clinically diagnosed with CdLS and showed mental and growth retardation. LH and FSH levels were low (FSH=0.7 UI/L, LH=0.5 UI/L), indicating hypogonadotropic hypogonadism. MRI identified small pituitary and the absence of olfactory bulbs, thus refining the additional diagnosis to KS.

Cornelia de Lange syndrome presents with variable phenotypic severity. *NIPBL* and *HDAC8* mutations typically underlie the most severe form of CdLS, while *SMC1A* and *SMC3* mutations are present in patients with mild, non-syndromic forms²²⁵. However, the broad spectrum of CdLS phenotypes makes it challenging to define clear genotype-phenotype distinctions. Therefore, all published reports describing the clinical aspects

of CdLS patients were reviewed to identify overlapping CHH phenotypes. The absence of menarche, lack of breast development or irregular menstruation is found in 13%, 20%, and 50% of female CdLS patients, respectively²²⁶. Cryptorchidism, hypoplastic genitalia or micropenis is reported in 45-98%²²⁶⁻²²⁸, 57%, and 37%^{226,229} of male CdLS patients. Additional phenotypes observed in both genders include cleft palate (20%-59%)^{229,230}, hearing loss (45%-60%)^{228,229}, and clinodactyly (74%)²²⁹.

Personal communications with Endocrinologists in Europe established collaborations which identified additional CdLS patients with indications of delayed puberty and/or infertility. An additional male patient with a mild form of CdLS (moderate mental retardation, short stature, ear anomalies, small and upturned nose) was identified from the clinical practice of Dr. Duarte Pignatelli (University of Porto, Portugal). In addition to meeting the criteria for CdLS, this patient also had a complete absence of puberty, micropenis, and low sex hormones levels (LH=0.3 UI/L, FSH=0.5 UI/L). MRI showed an absence of the left olfactory bulb and reduced size of the right, resulting in a diagnosis of Kallmann syndrome. In addition, the patient had hearing loss, scoliosis, and clinodactyly. Subsequent WES identified a heterozygous *NIPBL* frameshift variant (p.Pro2761fs). This variant was not observed in controls from ExAC, 1000Genomes, ESP or CoLaus datasets. This variant was not present in the proband's father, and DNA was not available from the deceased mother. Given that mother had no reported CdLS phenotype, it is likely that this mutation is *de novo*, which would be consistent with previous reports of mutations in *NIPBL*²³¹. The variant lies within the last exon (47) of *NIPBL*, and is predicted to result in the truncation of the last 43 amino acids of the protein (Figure 27).

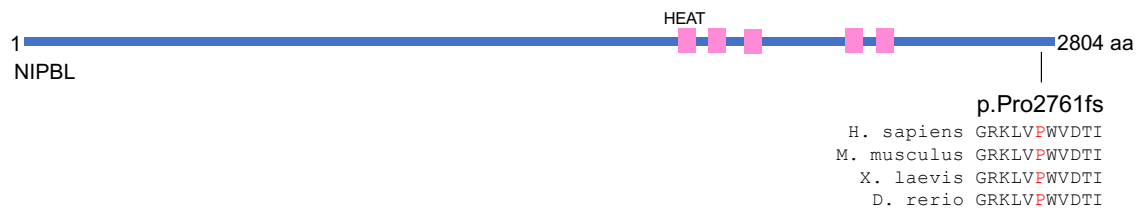


Figure 27. Schematic showing the p.Pro2761fs variant localized in NIPBL protein.

In summary, our data suggests potential phenotypic and genotypic overlap between CHH and CdLS. Based on this, all of the cohesin/CdLS genes were evaluated in our CHH cohort, however no additional rare, putatively pathogenic variants were identified. Additional investigation of the potential clinical and genetic overlap is ongoing.

Rare variant association studies in CHH patients vs. controls

A gene-collapsed burden test evaluates the number of individuals harboring a mutation in a given gene, and assumes all collapsed variants in a gene result in a deleterious effect. Variants in CHH European probands (n=159) and CoLaus (n=405) with MAF<1% were filtered to include PTVs, inframe InDels and missense changes predicted deleterious by SIFT and/or PolyPhen-2. Genome-wide significance ($p < 2.6 \times 10^{-6}$) was found for 17 genes, including *FGFR1* and *CHD7* — the two most frequently mutated genes in CHH (Figure 28).

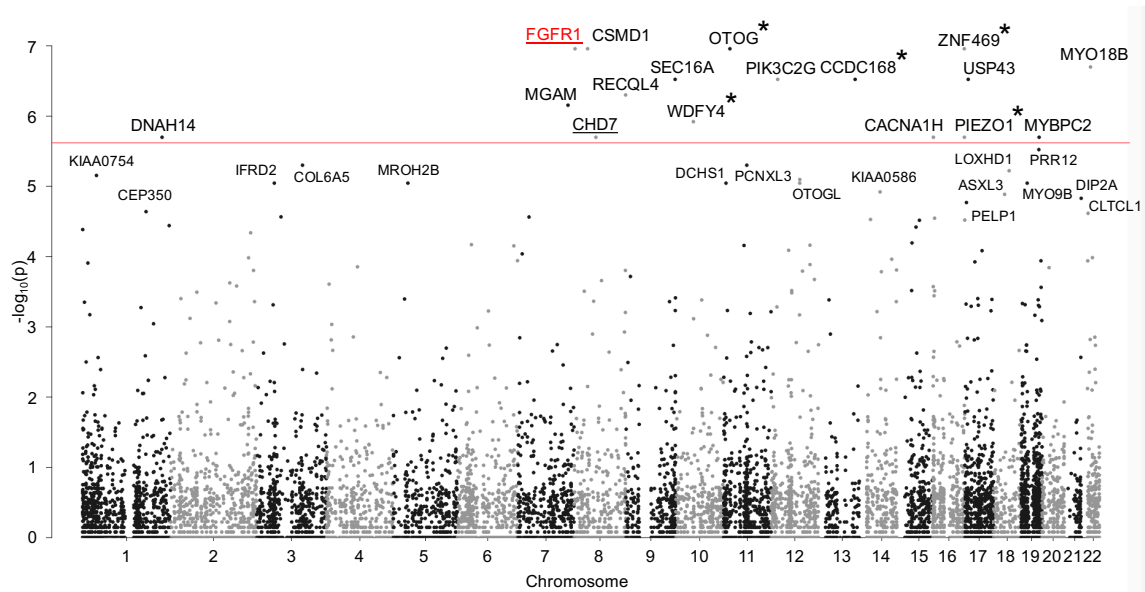


Figure 28. Gene-collapsed RVAS in CHH probands vs. CoLaus controls. Manhattan plot showing gene-collapsed associations in cases vs. controls. Red line defines genome-wide significance at $p=2.6 \times 10^{-6}$. Labeled genes below red line reached a marginal significance ($p=2.6 \times 10^{-5}$). Underlined genes are known CHH genes. Genes in red achieved genome-wide significance vs. both CoLaus and ExAC NFE controls. Genes in black achieved genome-wide significance vs. CoLaus and no association in ExAC EUR controls. Asterisks next to gene names indicate no coverage in ExAC individuals.

Gene-collapsed RVAS in CHH subgroups

Gene-collapsed burden analysis was then performed for the individual KS and nCHH subgroups. The analysis for KS ($n=91$) vs CoLaus identified 9 genes reaching genome-wide significance, including *FGFR1* and *CHD7* (Figure 29). The analysis for nCHH ($n=68$) vs. CoLaus demonstrated 4 genes (*WDFY4*, *ZNF469*, *FRMD4B*, and *DIP2A*) achieved genome-wide significance (Figure 30).

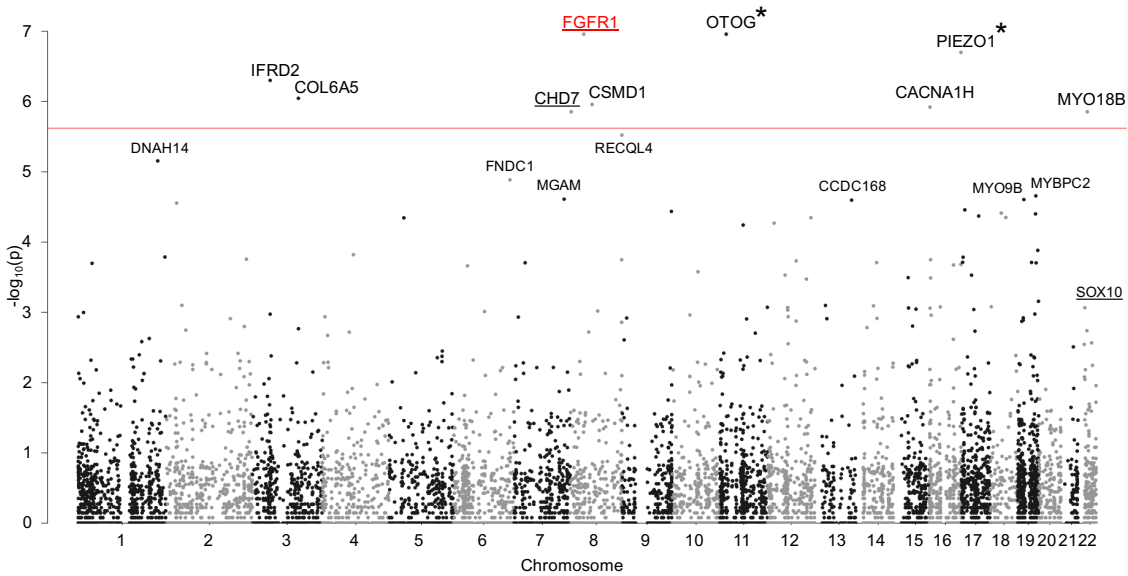


Figure 29. Gene-collapsed RVAS in KS probands vs. CoLaus controls. Manhattan plot showing gene-collapsed associations in cases vs. controls. Red line defines genome-wide significance at $p=2.6 \times 10^{-6}$. Labeled genes below red line reached a marginal significance ($p=2.6 \times 10^{-5}$). Underlined genes are known CHH genes. Genes in red achieved genome-wide significance vs. both CoLaus and ExAC NFE controls. Genes in black achieved genome-wide significance vs. CoLaus and no association in ExAC EUR controls. Asterisks next to gene names indicate no coverage in ExAC individuals.

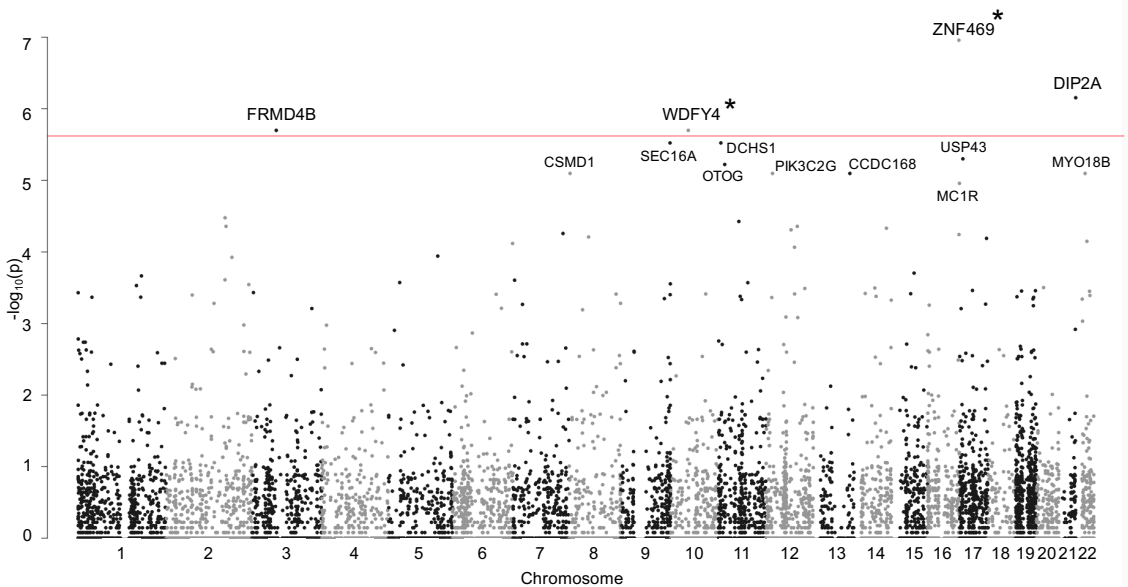


Figure 30. Gene-collapsed RVAS in nCHH probands vs. CoLaus controls. Manhattan plot showing gene-collapsed associations in cases vs. controls. Red line defines genome-wide significance at $p=2.6 \times 10^{-6}$. Labeled genes below red line reached a marginal significance ($p=2.6 \times 10^{-5}$). Underlined genes are known CHH genes. Genes in red achieved genome-wide significance vs. both CoLaus and ExAC NFE controls. Genes in black achieved genome-wide significance vs. CoLaus and no association in ExAC EUR controls. Asterisks next to gene names indicate no coverage in ExAC individuals.

This RVAS analysis examines the number of cases vs. controls harboring mutations in a given gene. However, the ExAC database is organized on a variant basis, thus it is not possible to determine if individuals contain multiple variants in the same gene. Using the raw number of variants present may lead to an over-estimation of the number of individuals harboring variants (e.g. two variants in one individual would be counted as two individuals). Thus, the ExAC database was used as a secondary analysis, despite its larger sample size. Of the 21 unique genes that were found to be significantly associated in the RVAS analysis using the CoLaus database, 6 genes (*COL6A5*, *WDFY4*, *OTOG*, *CCDC168*, *ZNF469*, and *PIEZO1*) were found to have low coverage (<10x) in the ExAC database and could not be evaluated. *FGFR1* was the only gene reaching genome-wide significance in the two controls datasets, while *USP43* achieved the closest significance to genome-wide threshold (Table 8).

Table 8. Coverage and RVAS results of genome-wide significant genes vs. ExAC NFE controls.

Gene	ExAC median coverage	RVAS P value
<i>FGFR1</i>	55	1.4E-12
<i>USP43</i>	32	6.0E-04
<i>DNAH14</i>	25	0.0014
<i>MYBPC2</i>	31	0.0055
<i>RECQL4</i>	30	0.0134
<i>CHD7</i>	42	0.0187
<i>IFRD2</i>	22	0.0191
<i>FRMD4B</i>	43	0.0403
<i>DIP2A</i>	43	0.0453
<i>PIK3C2G</i>	35	0.046
<i>CSMD1</i>	35	NS
<i>SEC16A</i>	40	NS
<i>MGAM</i>	47	NS
<i>CACNA1H</i>	26	NS
<i>MYO18B</i>	38	NS
<i>COL6A5</i>	9	-
<i>WDFY4</i>	0.003	-
<i>OTOG</i>	0.03	-
<i>CCDC168</i>	0	-
<i>ZNF469</i>	0.02	-
<i>PIEZO1</i>	0.2	-

Despite not being able to evaluate the complete set of significantly associated genes from the CoLaus analysis in the ExAC cohort due to low coverage, there are still compelling results that warrant additional follow-up. It will be a natural extension of this project to couple additional sample-centered control databases (when available) with an expanded CHH cohort to re-evaluate this promising analysis.

Discussion

Since the advent of HTS, hundreds of disease-causing genes have been identified in rare diseases²³². In most cases, unbiased family-based analyses showed high success rates in identifying mutations underlying disease phenotypes. Disease gene discovery for disorders related to puberty have benefited from using WES. Using stringent criteria (MAF<0.01% in public controls databases), family-based analysis of probands with central precocious puberty identified causative mutations in *MKRN3*²³³. Family-based recessive analyses using WES data also identified *SEMA3E*²³⁴, *POLR3B*²¹⁸, *FEZF1*⁵⁶, *RNF216* and *OTUD4*⁷¹ as novel genes implicated in CHH.

For the family-based analysis performed in this thesis, putatively pathogenic variants were selected by applying a stringent MAF cutoff (<0.1% for recessive, <0.01% for *de novo*). In fact, a recent retrospective analysis using well-characterized pathogenic mutations in the ClinVar and HGMD (Human Gene Mutation Database) databases from 60,706 individuals showed that the majority of mutations were indeed private or present in only one additional individual¹¹². Additionally, none of these mutations were present at MAF>0.1% in the ExAC population. Therefore, the classic cutoff to define rare variants (MAF<1%) is outdated.

The family-based analysis in 23 families led to the identification of new candidate genes underlying CHH. A homozygous mutation in *SLIT2* has been identified in a KS patient using an autosomal recessive analysis. *SLIT2* is a critical gene for the migration of neurons including GnRH neurons, retinal ganglion cells, and cerebellar granule neurons^{215,235,236}. *SLIT2* also interacts with the newly identified CHH genes *DCC*

and *NTN1* (Section 3) to directly control the migration of the neurons of the corpus callosum²³⁷.

Slit2 knockout mice show abnormal morphology of Purkinje cells dendrites in the cerebellum²³⁸, while *Slit1/Slit2* double knockouts show abnormal olfactory tract morphology — a structure critical for migrating of GnRH neurons during embryonic development²³⁹.

In addition to regulating neuronal migration, *SLIT2* has additional non-neurological functions. A recent study showed beige fat cells from adipose tissue secrete a cleaved fragment of SLIT2 (SLIT-C) which regulates metabolic functions and glucose homeostasis²⁴⁰. Interestingly, the patient harboring a *SLIT2* homozygous variant has a syndromic phenotype (i.e., metabolic defects, Dandy-Walker malformation, cerebellar ataxia, myopia and KS) which supports the known biology of *SLIT2*. This gene has been identified using an unbiased method such as family-based analysis. Additionally, a heterozygous *de novo* *MGAT1* variant was observed in a sporadic KS patient. *MGAT1* is involved in the homeostasis of gonadotropins by transferring and modifying N-acetylglucosamine residues onto glycoproteins such as FSH, LH, and hCG (reviewed in Bousfield and Dias²⁴¹). No human disease has been associated with *MGAT1* mutations, however mouse studies provide additional insight into the potential role of this gene in reproductive biology. Although *Mgat1*^{-/-} embryos die during early development (E10) and show critical defects in neural tube morphology,²⁴² studies on heterozygous knockout mice are particularly interesting. *Mgat1*^{+/-} male mice are poor breeders, while *Mgat1*^{+/-} female mice have reduced fertility due to a reduced number of

mature oocytes²⁴³, suggestive of a peripheral role in regulating fertility. Therefore, it is possible that *MGAT1* might exert both central and peripheral effects. Interestingly, the KS patient harboring a heterozygous *de novo* *MGAT1* variant was found to be resistant to FSH/hCG treatment which suggests a peripheral component to his reproductive phenotype.

A *de novo* *SMC3* missense variant was identified in a KS patient with a mild form of CdLS, a neurodevelopmental disorder characterized by mental and growth retardation, craniofacial defects and limb malformations. Indeed, *SMC3* mutations have been recently implicated with a, milder form of CdLS²²³. Notably, CHH-associated reproductive abnormalities (e.g., cryptorchidism, micropenis, absent menarche) and non-reproductive abnormalities (e.g., hearing loss, clinodactyly, cleft lip/palate) are also seen in CdLS patients. However, a precise diagnosis of hypogonadotropic hypogonadism is often not assessed in these patients given the severity of the primary neurological defects and the fact that puberty/fertility is inconsequential to both physician and CdLS patient families.

Expression patterns of *Smc3* in mouse are consistent with a role in regulating GnRH neuron ontogeny and migration. *Smc3*^{+/-} mice display developmental delay and craniofacial abnormalities²⁴⁴, the latter often associated with defects in the development of the olfactory placode, the origin of GnRH neurons²⁴⁵. A new project in our group has been recently established. The CRISPR/Cas9 technology is being employed with zebrafish having GFP-positive GnRH neurons to evaluate the effects of the *SMC3* p.Cys549Tyr on its orthologous gene counterpart.

In vitro studies have been performed to evaluate the functional effect of the p.Cys549Tyr variant on SMC3/SMC1 hinge binding. We showed the binding of mutated SMC3 hinge occurs and is more stable than wild-type, suggestive of a gain-of-function. Although counterintuitive, this gain may actually result in an impaired function of the cohesin complex. After the binding of the two SMC3 and SMC1 hinge domains to surround DNA fibers during gene transcription, the detachment of the complex is critical. The gain-of-function may result in the inability for the detachment of the cohesion complex, thus impacting the timing and/or coordination of additional regulatory functions.

Cohesin is ubiquitous, and has several other critical functions in the cell including 1) the maintenance of cohesive sister chromatids before cell divisions (i.e., chromosome biorientation)²⁴⁶, 2) genome compartmentalization through the organization of topologically-associated domains (TADs),²⁴⁷ and 3) regulation of transcription through physical interactions between transcription factors and distant promoters/enhancers²⁴⁸.

Human diseases caused by mutations in cohesin complex genes are usually defined as “cohesinopathies”, however given its growing importance in transcriptional regulation it has been suggested to change the terminology to “transcriptomopathies”.²⁴⁹ Interestingly, other known CHH genes identified within overlapping syndromes (e.g., *CHD7* in CHARGE syndrome⁶⁶, *SOX10* in Waardenburg syndrome⁶⁹) have also key functions in the transcriptional regulation during early development. After reaching out to the European network of GnRH deficiency investigators, we identified three additional CdLS patients with overlapping phenotypes. One CdLS patient was

additionally diagnosed with KS, and two with panhypopituitarism — a syndromic deficiency of pituitary hormones due to malformations in the hypothalamus, pituitary or their surrounding structures. WES in the second KS/CdLS patient revealed a putative *de novo* frameshift variant in the *NIPBL* gene. Heterozygous mutations in *NIPBL* are found in 60% of CdLS patients, and are often severe truncating *de novo* mutations which are subjected to nonsense mediated decay (NMD). As such, they are typically associated with the severe form of the disorder^{250,251}. However, our patient exhibited a mild form of CdLS, possibly due to the position of the frameshift variant in the last exon of *NIPBL*. Truncating mutations occurring in the last exon of genes usually escape NMD^{149,150}, therefore the patient may have the milder phenotype due to the production of a truncated (43 amino acid shorter) protein with attenuated function.

Combining information from the *SMC3* and *NIPBL* patients, it can be hypothesized that genes within cohesin complex play a role in the transcriptional regulation of key genes for GnRH biology. Several lines of evidences support this hypothesis. Human cell lines overexpressing *SMC3* show an upregulation of the transcription of many genes, including *GNRH1* and *GDF9*²⁵², an autocrine inducer of FSH production directly regulated by GnRH²⁵³. Zebrafish *nipblb* knockouts show neural defects and downregulation of the Wnt pathway²⁵⁴, crucial for the development of olfactory and forebrain regions supporting GnRH neuron migration²⁵⁵. Mutations in *SMC1A* also cause CdLS. Transcriptomic studies in cells derived from *SMC1A*-mutated CdLS patients showed dysregulation of the Notch pathway which is important for GnRH neuron maturation, migration, and transcriptional regulation^{256,257}.

Combined, these results suggest a phenotypic and genetic overlap between CdLS and CHH. However, to confirm an association of *SMC3* or other cohesin genes with isolated CHH, additional patients with mutations in this pathway will need to be uncovered.

Array-CGH was performed to identify putative pathogenic CNVs in a subset of CHH trios. We did not identify any compelling variants in the probands tested. A larger aCGH study previously conducted by Prof. Pitteloud did not observe any putative disease-causing CNVs in >70 CHH probands (data not published). Taken together, these data suggest at most a marginal role of large CNVs in CHH pathogenesis. However, a role of other structural variants cannot be excluded, as aCGH does not detect smaller CNVs (<20 kb), translocations, inversions, or other complex chromosomal aberrations. Other methods with higher resolution (depth-of-coverage from WES reads, WGS) are desired to better assess the mutational landscape of structural variation in CHH.

I performed an unbiased population-based rare-variant association study (RVAS) in CHH patients vs. controls. While GWAS are typically used to detect associations with common traits, RVAS recently emerged to identify associations of rare variants with rare diseases. This method also gains additional power by collapsing variants within distinct loci such as genes, regulatory regions, or pathways²⁰⁰. I identified several genes with significant genome-wide associations which will be further characterized. Genome-wide significance was achieved in only 2/34 known CHH genes — *FGFR1* and *CHD7*. As shown in Section 2, these are the most frequently mutated genes in CHH, with a prevalence of 15% and 13%, respectively. Nominal significance was observed for *SOX10*. Other CHH genes, including those with high

prevalence in CHH subgroups (e.g., *PROKR2* mutated in 8% of KS patients, *GNRHR* mutated in 7% of nCHH patients) did not yield significant associations. This may be due to a variety of reasons: 1) the sample size of cases/controls cohort is too small to detect additional associations, 2) confounding factors such as the presence of founder mutations in controls (e.g., *PROKR2* p.Leu173Arg, *TACR3* p.Trp275X) or carriers for recessive genes (e.g., *GNRHR*, *KISS1R*), or 3) some genes have a low prevalence even in CHH patients, indicating RVAS is not suitable for the identification of new genes with small or modest effects in the CHH population. Genes showing genome-wide significance and which were well-covered will be further characterized including evaluating variant effects, segregation in affected relatives, expression in GnRH-related tissues.

Additionally, a unidirectional RVAS burden analysis was performed, meaning the algorithm considers all collapsed variants for the gene to result in deleterious effects. This test has been shown to be successful for gene discovery in diseases known to be primarily caused by loss-of-function mutations^{258,259}. However, this test can underperform when complex mechanisms such as varying magnitudes of loss-of-function mutations, gain-of-function mutations, or oligogenicity underlie the phenotype²⁶⁰. Other RVAS methods such as variance-component tests (e.g., SKAT²⁶¹, C-alpha²⁶², SSU²⁶³) evaluate potential enrichment of rare variants in cases vs. controls, regardless of protective or deleterious effects. Such tests take into account the overall distribution of rare variants in cases vs. controls, aggregating the single variant scores into a unique signal. For this reason, variance-component tests are suited for complex diseases, where phenotype traits are dependent on the magnitude of variant

deleterious effects, or where both loss-of-function and gain-of-function rare variants result in disease phenotype.^{264,265} Further, to overcome the limitations of both linear and mixed models, combined tests such as SKAT-O²⁶⁶ and MiST²⁶⁷ have been developed. Such alternative tests could be used in CHH cases vs. controls to overcome the limitations seen in the burden tests used in this thesis.

In summary, both the family-based and population-based analyses generated promising candidate genes associated with CHH, and underscore the need for a multi-faceted approach to identify genes related to diseases with complex genotypes and phenotypes.

Thesis Conclusions and Future perspective

HTS has dramatically increased the success rates in identifying disease-causing mutations in human disorders, especially those which are rare in the population²⁶⁸. I applied HTS and bioinformatics to clarify the genetic architecture and discover new genes associated with CHH, a rare disease characterized by absent puberty and infertility due to GnRH deficiency. In Section 1, a bioinformatics pipeline was implemented to process >1,000 exomes, generating high quality, reliable data for further genetic studies. This pipeline is now routinely used in the Pitteloud laboratory, and can be considered an important deliverable of this thesis project and for future projects within the department. This pipeline will be further improved in the future based upon the availability of new bioinformatics tools to guarantee faster analyses and accurate data.

One limitation of this thesis study design is that the putative pathogenic effects of the identified variants – especially in the upstream filtering processes – are not assessed experimentally (i.e., *in vitro* or *in vivo*). In fact, the two major factors to define a putative pathogenic variant were 1) allele frequency in the general population and 2) *in silico* predictions of its effects on protein functionality. As described earlier, the MAF cutoff used for new gene discovery was defined using a retrospective study in thousands of individuals. In fact, all well-characterized pathogenic variants reported in Mendelian disorders had an $MAF < 0.1\%$ ¹¹². However, one cannot exclude *a priori* that one or more variants with an MAF above the define threshold would still cause – or

participate – in the manifestation of the disease phenotype(s). The potential deleterious effects of variants – especially the frequently observed missense ones – have been assessed using the two most commonly used and popular algorithms, SIFT¹⁴⁵ and PolyPhen-2¹⁴⁶. The combination of these two programs was intended to complement their algorithms, as SIFT predicts variants deleteriousness based on sequence homology among several species, while PolyPhen-2 mainly predicts deleterious variants using protein structure information. Despite their well-known sensitivity (i.e., correct identification of true positives), all algorithms – including SIFT and PolyPhen-2 – suffer of low specificity^{269,270}. Although the

Before embarking on a new gene discovery, it was imperative to characterize the known genetic architecture of CHH and to identify probands harboring mutations which already explain their phenotype. Section 2 describes a comprehensive screening of the known CHH genes. CHH is a heterogeneous disease both clinically and genetically, and oligogenicity in CHH patients has been already reported. With this thesis, I show that oligogenic inheritance in CHH patients is more common than previously thought. Thus, CHH displays a dual genetic pattern — a Mendelian, monogenic inheritance (e.g., *de novo* variants in *FGFR1*, homozygous variants in *TACR3*, hemizygous variants in *ANOS1*), as well as a non-Mendelian, oligogenic inheritance. WES demonstrated that half of CHH patients harbor at least one putatively pathogenic variant in a known gene — a higher frequency than observed in previous studies. I analyzed for the first time the full set of

known genes underlying both CHH and CDGP. No genetic overlap – screening the known implicated genes so far – between the two forms of GnRH deficiency was observed. However, this result is of critical importance for future studies aiming to develop targeted genetic testing to correctly diagnose CHH in adolescent patients presenting with delayed puberty.

To identify new genes implicated in CHH, differing strategies were employed to tackle the main scientific question. First, the known information regarding GnRH biology and CHH pathogenesis was exploited to target proteins with domains that are recurrently present in known CHH genes and participate in neuronal migration. Loss-of-function mutations in *DCC* and its ligand *NTN1* were detected in CHH patients. In a collaborative effort within the Pitteloud lab and external collaborators, we demonstrated the critical role of DCC and Netrin-1 in GnRH migration during early development, implicating these two genes in CHH pathogenesis. Similar parallel projects are ongoing within the Pitteloud group using the strategy of targeting other recurrent protein domains encoded by known CHH genes.

Unbiased strategies were performed to identify new genes implicated in CHH. Family-based analysis targeting *de novo*, autosomal recessive and X-linked recessive inherited variants in CHH probands identified several promising candidates. These will now be further characterized using *in vivo* and *in vitro* studies collaboratively conducted by other members of Pitteloud group. Specifically, I report a patient with the dual diagnosis of CdLS and CHH harboring a *de novo* variant in the *SMC3* gene, a member of the cohesin complex. *In silico* and *in vitro* experiments demonstrated impaired function of mutant SMC3. A second patient with CHH and CdLS has been

identified harboring mutation in *NIPBL*, another CdLS gene. Several lines of evidence suggest a clinical overlap between CdLS and CHH, thus suggesting a role for CdLS genes in the pathophysiology of CHH through the disruption of critical pathways in key organs for GnRH biology during early development.

Phenotypic overlap is common among neurodevelopmental disorders, especially in syndromic forms. However, the results shown in this thesis suggest that the phenotypic overlap between CHH and CdLS might be also mirrored by partially shared genetic patterns. Genetic overlap between syndromes has been reported previously²⁷¹, and a recent study showed that 10% of patients with Smith-Magenis syndrome harbored pathogenic mutations in genes known to implicate other syndromes with overlapping phenotypes (e.g., Kabuki syndrome, autism spectrum disorder)²⁷². Such overlap can be explained by many factors, acting alone or in combination. Mutations in these genes participate in same or similar pathways that – if perturbed – result in a disease phenotype. In the specific case of this thesis work, transcriptional regulation is the most plausible pathway to be disrupted in CHH and CdLS patients. Although one cannot exclude the presence of false positives due to misdiagnosis, it is highly unlikely to be the case for all instances of overlapping phenotypes among different disorders. One possible explanation is that – depending on the type of mutation – each gene can be implicated in more than one disorder. Two striking examples are *CHD7* and *SOX10*, where usually PTVs underlie more severe phenotypes such as CHARGE and Waardenburg syndromes, while missense variants are present in CHH patients^{69,101}. Depending by other intervening factors (i.e. additional mutations in other genes,

environmental factors, different epigenetic signatures) similar mutations in the same gene can result in highly variable phenotypes.

A second unbiased strategy employed a statistical analysis (RVAS) to observe a cumulative, significant enrichment of rare variants in cases compared to controls. As a proof-of-principle of the validity of gene-collapsed RVAS, genome-wide significant associations were achieved by the two most frequently mutated genes in CHH patients. However, given the small sample size of our cases and controls cohorts, the statistical power of this analysis was not sufficient to generate genome-wide significance for other CHH genes. The release of larger sample-centered datasets of controls with available genotype data, together with an expanded CHH cohort will increase the statistical power of the RVAS analysis. A pathway-collapsed RVAS strategy is already part of an ongoing collaboration with Zoltán Kutalik, PhD (Institute of Social and Preventive Medicine, CHUV).

We are witnessing an unprecedented era for human genetics. The rapid advancement of sequencing technologies, together with the development of fast, reliable and accurate bioinformatics tools is permitting to finally shed light on the complex molecular links between genotypes and phenotypes. The advent of HTS and WES in particular expanded our knowledge on human variation, increasing the number of individuals in control datasets by almost two orders of magnitude in few years. This thesis work started in 2012, when the reference dataset for human variation was the 1000 Genomes Project, providing low-coverage, whole-genome genetic information on 1,092 controls from different populations¹¹⁵. Later, the number of sequenced controls in the 1000 Genomes Project raised to 2,504²⁷³. Other parallel projects using WES in

thousands of individuals became available to the public, such as the NHLBI Exome Sequencing Project (ESP) with 6,500 exomes (<http://evs.gs.washington.edu/EVS/>) and the Exome Aggregation Consortium (ExAC), with 66,076 exomes ¹¹². At the present time, the largest public database of human variation is the Genome Aggregation Database (gnomAD), with ~120,000 exomes and ~14,000 genomes sequenced (<http://gnomad.broadinstitute.org>). Given the increasing availability of sequencing methods and the ongoing efforts of private and public entities to sequence as many as possible persons, it is estimated that up to 2 billion of human genomes could be sequenced by 2025²⁷⁴.

Thus, the growing availability of genetic tests raises the need for collaboration of many stakeholders and professionals: 1) computer scientists and services for data storage and computational analyses, 2) bioinformaticians and geneticists to process and interpret the information deriving from such data, and 3) clinicians to deliver genetic results to patients – in collaboration with genetic counsellors. Such multi-disciplinary effort is desired in a society where the “uberization” of medicine²⁷⁵ is spreading in the population, in order to tackle any dangerous outcome (i.e., self-diagnosis or self-medication in patients receiving a genetic testing).

With this thesis, I demonstrate the exceptional power of WES coupled with bioinformatics to increase our knowledge on the genetic architecture of CHH, and to identify with a faster pace new genes associated with the disease. However, we are only scratching the surface. In the last years, WES has been the technology of choice over whole-genome sequencing (WGS) for several reasons: 1) it was 3-4 times cheaper than WGS, 2) it generated a higher throughput of reads compared to WGS, and 3) it covered

the exons of protein-coding genes, where 85% of disease-causing variants have been reported to date^{29,30}. The dream of the \$1,000 genome has lately become reality, and recent claims from sequencing machine companies promise to bring the price of a full genome sequence to \$100 after 2018 (<https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2236383>). With these premises, the “democratization” of genome analysis²⁷⁶ and of healthcare in general²⁷⁷ is not far off. WGS will soon replace WES to investigate genomic variation—a change we are already beginning to see today.

In this thesis, WES and aCGH were performed to identify putative pathogenic SNVs, InDels and CNVs in CHH patients. However, the future implementation of WGS will enable researchers to more comprehensively investigate the genetic etiology of CHH. This will include the ability to evaluate additional types of variation in non-coding regions, regions regulating the structural organization of the genome, and complex chromosomal abnormalities not currently detected by traditional WES. The roles of such variation in disease pathogenesis has only recently been highlighted, and the interest in these fields is growing²⁷⁸⁻²⁸¹. WGS will be the all-in-one experiment to evaluate the full picture of genomic variation in CHH patients, and will result in a better knowledge of the genetic etiology of a spectrum of GnRH deficiency disorders.

References

- 1 Lejeune, J., Gautier, M. & Turpin, R. Etude Des Chromosomes Somatiques De Neuf Enfants Mongoliens. *Cr Hebd Acad Sci* **248**, 1721-1722 (1959).
- 2 Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *J Mol Diagn* **6**, 411-411 (2004).
- 3 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**, 441-448 (1975).
- 4 Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073-1080 (1989).
- 5 Koenig, M. *et al.* Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50**, 509-517 (1987).
- 6 Wallace, M. R. *et al.* Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249**, 181-186 (1990).
- 7 Moore, K. J. Utilization of mouse models in the discovery of human disease genes. *Drug discovery today* **4**, 123-128 (1999).
- 8 Tabor, H. K., Risch, N. J. & Myers, R. M. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature reviews. Genetics* **3**, 391-397, doi:10.1038/nrg796 (2002).
- 9 Dode, C. *et al.* Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat Genet* **33**, 463-465, doi:10.1038/ng1122 (2003).
- 10 Muenke, M. *et al.* A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet* **8**, 269-274, doi:10.1038/ng1194-269 (1994).
- 11 Rossi, M., Jones, R. L., Norbury, G., Bloch-Zupan, A. & Winter, R. M. The appearance of the feet in Pfeiffer syndrome caused by FGFR1 P252R mutation. *Clin Dysmorphol* **12**, 269-274, doi:10.1097/01.mcd.0000089130.98648.21 (2003).
- 12 White, K. E. *et al.* Mutations that cause osteoglophonic dysplasia define novel roles for FGFR1 in bone elongation. *Am J Hum Genet* **76**, 361-367, doi:10.1086/427956 (2005).
- 13 Sow, A. J. *et al.* Osteoglophonic dysplasia: A 'common' mutation in a rare disease. *Clin Genet* **78**, 197-198, doi:10.1111/j.1399-0004.2010.01382.x (2010).
- 14 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).
- 15 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).
- 16 International HapMap, C. A haplotype map of the human genome. *Nature* **437**, 1299-1320, doi:10.1038/nature04226 (2005).
- 17 Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).
- 18 Cohen, J. C., Boerwinkle, E., Mosley, T. H., Jr. & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* **354**, 1264-1272, doi:10.1056/NEJMoa054013 (2006).

- 19 Kotowski, I. K. *et al.* A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* **78**, 410-422, doi:10.1086/500615 (2006).
- 20 Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-9367, doi:10.1073/pnas.0903103106 (2009).
- 21 Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am J Hum Genet* **90**, 7-24, doi:10.1016/j.ajhg.2011.11.029 (2012).
- 22 Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8, doi:10.1016/j.ygeno.2015.11.003 (2016).
- 23 Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A. P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* **36**, e25, doi:10.1093/nar/gkn021 (2008).
- 24 Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet* **59**, 5-15, doi:10.1038/jhg.2013.114 (2014).
- 25 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**, 30-35, doi:10.1038/ng.499 (2010).
- 26 Wetterstrand, K. A. *DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP)*. <<http://www.genome.gov/sequencingcosts>> (2013).
- 27 Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews. Genetics* **12**, 745-755, doi:10.1038/nrg3031 (2011).
- 28 Meynert, A. M., Ansari, M., FitzPatrick, D. R. & Taylor, M. S. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC bioinformatics* **15**, 247, doi:10.1186/1471-2105-15-247 (2014).
- 29 Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* **33 Suppl**, 228-237, doi:10.1038/ng1090 (2003).
- 30 Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. What can exome sequencing do for you? *J Med Genet* **48**, 580-589, doi:10.1136/jmedgenet-2011-100223 (2011).
- 31 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 32 Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A* **112**, 5473-5478, doi:10.1073/pnas.1418631112 (2015).
- 33 Beales, P. L. *et al.* Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am J Hum Genet* **72**, 1187-1199, doi:10.1086/375178 (2003).
- 34 Kajiwara, K., Berson, E. L. & Dryja, T. P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-1608 (1994).
- 35 van Blitterswijk, M. *et al.* Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet* **21**, 3776-3784, doi:10.1093/hmg/dds199 (2012).
- 36 Gabriel, S. B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* **31**, 89-93, doi:10.1038/ng868 (2002).

- 37 Priest, J. R. *et al.* De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet* **12**, e1005963, doi:10.1371/journal.pgen.1005963 (2016).
- 38 Gazzo, A. M. *et al.* DIDA: A curated and annotated digenic diseases database. *Nucleic Acids Res* **44**, D900-907, doi:10.1093/nar/gkv1068 (2016).
- 39 Sykiotis, G. P. *et al.* Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. *Proc Natl Acad Sci U S A* **107**, 15140-15144, doi:10.1073/pnas.1009622107 (2010).
- 40 Antunes, J. L., Carmel, P. W., Housepian, E. M. & Ferin, M. Luteinizing hormone-releasing hormone in human pituitary blood. *Journal of neurosurgery* **49**, 382-386, doi:10.3171/jns.1978.49.3.0382 (1978).
- 41 Terasawa, E. Luteinizing hormone-releasing hormone (LHRH) neurons: mechanism of pulsatile LHRH release. *Vitamins and hormones* **63**, 91-129 (2001).
- 42 Wray, S., Grant, P. & Gainer, H. Evidence that cells expressing luteinizing hormone-releasing hormone mRNA in the mouse are derived from progenitor cells in the olfactory placode. *Proc Natl Acad Sci U S A* **86**, 8132-8136 (1989).
- 43 Seminara, S. B., Hayes, F. J. & Crowley, W. F., Jr. Gonadotropin-releasing hormone deficiency in the human (idiopathic hypogonadotropic hypogonadism and Kallmann's syndrome): pathophysiological and genetic considerations. *Endocrine reviews* **19**, 521-539, doi:10.1210/edrv.19.5.0344 (1998).
- 44 Kim, H. G. *et al.* Hypogonadotropic hypogonadism and cleft lip and palate caused by a balanced translocation producing haploinsufficiency for FGFR1. *J Med Genet* **42**, 666-672, doi:10.1136/jmg.2004.026989 (2005).
- 45 Chan, Y. M. *et al.* GNRH1 mutations in patients with idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci U S A* **106**, 11703-11708, doi:10.1073/pnas.0903449106 (2009).
- 46 Laitinen, E. M., Hero, M., Vaaralahti, K., Tommiska, J. & Raivio, T. Bone mineral density, body composition and bone turnover in patients with congenital hypogonadotropic hypogonadism. *International journal of andrology* **35**, 534-540, doi:10.1111/j.1365-2605.2011.01237.x (2012).
- 47 Villanueva, C. *et al.* Congenital hypogonadotropic hypogonadism with split hand/foot malformation: a clinical entity with a high frequency of FGFR1 mutations. *Genetics in medicine : official journal of the American College of Medical Genetics*, doi:10.1038/gim.2014.166 (2014).
- 48 Quinton, R. *et al.* The neuroradiology of Kallmann's syndrome: a genotypic and phenotypic analysis. *The Journal of clinical endocrinology and metabolism* **81**, 3010-3017, doi:10.1210/jcem.81.8.8768867 (1996).
- 49 Coatesworth, A. P. & Woodhead, C. J. Conductive hearing loss associated with Kallmann's syndrome. *The Journal of laryngology and otology* **116**, 125-126 (2002).
- 50 Kirk, J. M. *et al.* Unilateral renal aplasia in X-linked Kallmann's syndrome. *Clin Genet* **46**, 260-262 (1994).
- 51 Pitteloud, N. *et al.* Mutations in fibroblast growth factor receptor 1 cause Kallmann syndrome with a wide spectrum of reproductive phenotypes. *Molecular and cellular endocrinology* **254-255**, 60-69, doi:10.1016/j.mce.2006.04.021 (2006).

- 52 Pitteloud, N., Durrani, S., Raivio, T. & Sykiotis, G. P. Complex genetics in idiopathic hypogonadotropic hypogonadism. *Frontiers of hormone research* **39**, 142-153, doi:10.1159/000312700 (2010).
- 53 Layman, L. C. The genetic basis of female reproductive disorders: etiology and clinical testing. *Molecular and cellular endocrinology* **370**, 138-148, doi:10.1016/j.mce.2013.02.016 (2013).
- 54 Topaloglu, A. K. *et al.* Inactivating KISS1 mutation and hypogonadotropic hypogonadism. *N Engl J Med* **366**, 629-635, doi:10.1056/NEJMoa1111184 (2012).
- 55 Topaloglu, A. K. *et al.* TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat Genet* **41**, 354-358, doi:10.1038/ng.306 (2009).
- 56 Kotan, L. D. *et al.* Mutations in FEZF1 cause Kallmann syndrome. *Am J Hum Genet* **95**, 326-331, doi:10.1016/j.ajhg.2014.08.006 (2014).
- 57 Franco, B. *et al.* A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* **353**, 529-536, doi:10.1038/353529a0 (1991).
- 58 Kim, H. G. *et al.* WDR11, a WD protein that interacts with transcription factor EMX1, is mutated in idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. *Am J Hum Genet* **87**, 465-479, doi:10.1016/j.ajhg.2010.08.018 (2010).
- 59 Tornberg, J. *et al.* Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci U S A* **108**, 11524-11529, doi:10.1073/pnas.1102284108 (2011).
- 60 Miraoui, H. *et al.* Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. *Am J Hum Genet* **92**, 725-743, doi:10.1016/j.ajhg.2013.04.008 (2013).
- 61 Kansakoski, J. *et al.* Mutation screening of SEMA3A and SEMA7A in patients with congenital hypogonadotropic hypogonadism. *Pediatric research* **75**, 641-644, doi:10.1038/pr.2014.23 (2014).
- 62 Miura, K., Acierno, J. S., Jr. & Seminara, S. B. Characterization of the human nasal embryonic LHRH factor gene, NELF, and a mutation screening among 65 patients with idiopathic hypogonadotropic hypogonadism (IHH). *J Hum Genet* **49**, 265-268, doi:10.1007/s10038-004-0137-4 (2004).
- 63 de Roux, N. *et al.* A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *N Engl J Med* **337**, 1597-1602, doi:10.1056/NEJM199711273372205 (1997).
- 64 Dode, C. *et al.* Kallmann syndrome: mutations in the genes encoding prokineticin-2 and prokineticin receptor-2. *PLoS Genet* **2**, e175, doi:10.1371/journal.pgen.0020175 (2006).
- 65 Falardeau, J. *et al.* Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *J Clin Invest* **118**, 2822-2831, doi:10.1172/JCI34538 (2008).
- 66 Kim, H. G. *et al.* Mutations in CHD7, encoding a chromatin-remodeling protein, cause idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. *Am J Hum Genet* **83**, 511-519, doi:10.1016/j.ajhg.2008.09.005 (2008).

- 67 Bouligand, J. *et al.* Isolated familial hypogonadotropic hypogonadism and a GNRH1 mutation. *N Engl J Med* **360**, 2742-2748, doi:10.1056/NEJMoa0900136 (2009).
- 68 Young, J. *et al.* SEMA3A deletion in a family with Kallmann syndrome validates the role of semaphorin 3A in human puberty and olfactory system development. *Human reproduction* **27**, 1460-1465, doi:10.1093/humrep/des022 (2012).
- 69 Pingault, V. *et al.* Loss-of-function mutations in SOX10 cause Kallmann syndrome with deafness. *Am J Hum Genet* **92**, 707-724, doi:10.1016/j.ajhg.2013.03.024 (2013).
- 70 Salian-Mehta, S. *et al.* Functional consequences of AXL sequence variants in hypogonadotropic hypogonadism. *The Journal of clinical endocrinology and metabolism* **99**, 1452-1460, doi:10.1210/jc.2013-3426 (2014).
- 71 Margolin, D. H. *et al.* Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. *N Engl J Med* **368**, 1992-2003, doi:10.1056/NEJMoa1215993 (2013).
- 72 Boehm, U. *et al.* Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism--pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol* **11**, 547-564, doi:10.1038/nrendo.2015.112 (2015).
- 73 Kotan, L. D. *et al.* Idiopathic Hypogonadotropic Hypogonadism Caused by Inactivating Mutations in SRA1. *J Clin Res Pediatr Endocrinol* **8**, 125-134, doi:10.4274/jcrpe.3248 (2016).
- 74 Hanchate, N. K. *et al.* SEMA3A, a gene involved in axonal pathfinding, is mutated in patients with Kallmann syndrome. *PLoS Genet* **8**, e1002896, doi:10.1371/journal.pgen.1002896 (2012).
- 75 Gianetti, E. *et al.* TAC3/TACR3 mutations reveal preferential activation of gonadotropin-releasing hormone release by neurokinin B in neonatal life followed by reversal in adulthood. *The Journal of clinical endocrinology and metabolism* **95**, 2857-2867, doi:10.1210/jc.2009-2320 (2010).
- 76 Merke, D. P., Tajima, T., Baron, J. & Cutler, G. B., Jr. Hypogonadotropic hypogonadism in a female caused by an X-linked recessive mutation in the DAX1 gene. *N Engl J Med* **340**, 1248-1252, doi:10.1056/NEJM199904223401605 (1999).
- 77 Zenaty, D. *et al.* Paediatric phenotype of Kallmann syndrome due to mutations of fibroblast growth factor receptor 1 (FGFR1). *Molecular and cellular endocrinology* **254-255**, 78-83, doi:10.1016/j.mce.2006.04.006 (2006).
- 78 Miura, K. *et al.* A case of Kallmann syndrome carrying a missense mutation in alternatively spliced exon 8A encoding the immunoglobulin-like domain IIIb of fibroblast growth factor receptor 1. *Human reproduction* **25**, 1076-1080, doi:10.1093/humrep/deq006 (2010).
- 79 Jongmans, M. C. *et al.* CHD7 mutations in patients initially diagnosed with Kallmann syndrome--the clinical overlap with CHARGE syndrome. *Clin Genet* **75**, 65-71, doi:10.1111/j.1399-0004.2008.01107.x (2009).
- 80 Vaaralahti, K. *et al.* De novo SOX10 nonsense mutation in a patient with Kallmann syndrome and hearing loss. *Pediatric research* **76**, 115-116, doi:10.1038/pr.2014.60 (2014).
- 81 Suzuki, E. *et al.* De novo frameshift mutation in fibroblast growth factor 8 in a male patient with gonadotropin deficiency. *Hormone research in paediatrics* **81**, 139-144, doi:10.1159/000355380 (2014).

- 82 Pitteloud, N. *et al.* Loss-of-function mutation in the prokineticin 2 gene causes Kallmann syndrome and normosmic idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci U S A* **104**, 17447-17452, doi:10.1073/pnas.0707173104 (2007).
- 83 Seminara, S. B. *et al.* The GPR54 gene as a regulator of puberty. *N Engl J Med* **349**, 1614-1627, doi:10.1056/NEJMoa035322 (2003).
- 84 Costa-Barbosa, F. A. *et al.* Prioritizing genetic testing in patients with Kallmann syndrome using clinical phenotypes. *The Journal of clinical endocrinology and metabolism* **98**, E943-953, doi:10.1210/jc.2012-4116 (2013).
- 85 Villanueva, C. *et al.* Congenital hypogonadotropic hypogonadism with split hand/foot malformation: a clinical entity with a high frequency of FGFR1 mutations. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, 651-659, doi:10.1038/gim.2014.166 (2015).
- 86 Farooqi, I. S. & O'Rahilly, S. Mutations in ligands and receptors of the leptin-melanocortin pathway that lead to obesity. *Nat Clin Pract Endocrinol Metab* **4**, 569-577, doi:10.1038/ncpendmet0966 (2008).
- 87 Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet* **16**, 303-306, doi:10.1038/ng0797-303 (1997).
- 88 Read, A. P. & Newton, V. E. Waardenburg syndrome. *J Med Genet* **34**, 656-665 (1997).
- 89 Tamayo, M. L. *et al.* Screening program for Waardenburg syndrome in Colombia: clinical definition and phenotypic variability. *Am J Med Genet A* **146A**, 1026-1031, doi:10.1002/ajmg.a.32189 (2008).
- 90 Pusch, C. *et al.* The SOX10/Sox10 gene from human and mouse: sequence, expression, and transactivation by the encoded HMG domain transcription factor. *Hum Genet* **103**, 115-123 (1998).
- 91 Pingault, V. *et al.* SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* **18**, 171-173, doi:10.1038/ng0298-171 (1998).
- 92 Bondurand, N. *et al.* Deletions at the SOX10 gene locus cause Waardenburg syndrome types 2 and 4. *Am J Hum Genet* **81**, 1169-1185, doi:10.1086/522090 (2007).
- 93 Elmaleh-Berges, M. *et al.* Spectrum of temporal bone abnormalities in patients with Waardenburg syndrome and SOX10 mutations. *AJNR Am J Neuroradiol* **34**, 1257-1263, doi:10.3174/ajnr.A3367 (2013).
- 94 Vissers, L. E. *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* **36**, 955-957, doi:10.1038/ng1407 (2004).
- 95 Marfella, C. G. & Imbalzano, A. N. The Chd family of chromatin remodelers. *Mutat Res* **618**, 30-40, doi:10.1016/j.mrfmmm.2006.07.012 (2007).
- 96 Bajpai, R. *et al.* CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958-962, doi:10.1038/nature08733 (2010).
- 97 Wheeler, P. G., Quigley, C. A., Sadeghi-Nejad, A. & Weaver, D. D. Hypogonadism and CHARGE association. *Am J Med Genet* **94**, 228-231 (2000).
- 98 Chalouhi, C. *et al.* Olfactory evaluation in children: application to the CHARGE syndrome. *Pediatrics* **116**, e81-88, doi:10.1542/peds.2004-1970 (2005).
- 99 Pinto, G. *et al.* CHARGE syndrome includes hypogonadotropic hypogonadism and abnormal olfactory bulb development. *The Journal of clinical endocrinology and metabolism* **90**, 5621-5626, doi:10.1210/jc.2004-2474 (2005).

- 100 Bergman, J. E. *et al.* The results of CHD7 analysis in clinically well-characterized patients with Kallmann syndrome. *The Journal of clinical endocrinology and metabolism* **97**, E858-862, doi:10.1210/jc.2011-2652 (2012).
- 101 Marcos, S. *et al.* The prevalence of CHD7 missense versus truncating mutations is higher in patients with Kallmann syndrome than in typical CHARGE patients. *The Journal of clinical endocrinology and metabolism* **99**, E2138-2143, doi:10.1210/jc.2014-2110 (2014).
- 102 Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H. & Inoue, I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* **57**, 621-632, doi:10.1038/jhg.2012.91 (2012).
- 103 Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* **369**, 1502-1511, doi:10.1056/NEJMoa1306555 (2013).
- 104 Chong, J. X. S., J.; Nickerson, D.A.; Bamshad M.J. Assessment of the success rate of two years of large-scale exome sequencing efforts to identify genes for Mendelian conditions at the University of Washington Center for Mendelian Genomics. (2014).
- 105 Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Disease gene identification strategies for exome sequencing. *European journal of human genetics : EJHG* **20**, 490-497, doi:10.1038/ejhg.2011.258 (2012).
- 106 Stitzel, N. O., Kiezun, A. & Sunyaev, S. Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol* **12**, 227, doi:10.1186/gb-2011-12-9-227 (2011).
- 107 Pitteloud, N. *et al.* The role of prior pubertal development, biochemical markers of testicular maturation, and genetics in elucidating the phenotypic heterogeneity of idiopathic hypogonadotropic hypogonadism. *The Journal of clinical endocrinology and metabolism* **87**, 152-160, doi:10.1210/jcem.87.1.8131 (2002).
- 108 Lewkowitz-Shpuntoff, H. M. *et al.* Olfactory phenotypic spectrum in idiopathic hypogonadotropic hypogonadism: pathophysiological and genetic implications. *The Journal of clinical endocrinology and metabolism* **97**, E136-144, doi:10.1210/jc.2011-2041 (2012).
- 109 Palmert, M. R. & Dunkel, L. Clinical practice. Delayed puberty. *N Engl J Med* **366**, 443-453, doi:10.1056/NEJMcp1109290 (2012).
- 110 Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC cardiovascular disorders* **8**, 6, doi:10.1186/1471-2261-8-6 (2008).
- 111 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
- 112 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
- 113 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 114 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).
- 115 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

- 116 Mills, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome research* **16**, 1182-1190, doi:10.1101/gr.4565806 (2006).
- 117 Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research* **8**, 175-185 (1998).
- 118 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research* **8**, 186-194 (1998).
- 119 International HapMap, C. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58, doi:10.1038/nature09298 (2010).
- 120 The 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).
- 121 Bainbridge, M. N. *et al.* Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. *Genome Biol* **12**, R68, doi:10.1186/gb-2011-12-7-r68 (2011).
- 122 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).
- 123 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 124 Garrison, E. M., G. Haplotype-based variant detection from short-read sequencing. *arXiv* (2012).
- 125 Hwang, S., Kim, E., Lee, I. & Marcotte, E. M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep* **5**, 17875, doi:10.1038/srep17875 (2015).
- 126 Wang, J. *et al.* Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. *Sci Rep* **6**, 33256, doi:10.1038/srep33256 (2016).
- 127 Yu, X. & Sun, S. Comparing a few SNP calling algorithms using low-coverage sequencing data. *BMC bioinformatics* **14**, 274, doi:10.1186/1471-2105-14-274 (2013).
- 128 Meienberg, J. *et al.* New insights into the performance of human whole-exome capture platforms. *Nucleic Acids Res* **43**, e76, doi:10.1093/nar/gkv216 (2015).
- 129 Gajdos, Z. K., Hirschhorn, J. N. & Palmert, M. R. What controls the timing of puberty? An update on progress from genetic investigation. *Curr Opin Endocrinol Diabetes Obes* **16**, 16-24, doi:10.1097/MED.0b013e328320253c (2009).
- 130 Zhu, J. *et al.* A shared genetic basis for self-limited delayed puberty and idiopathic hypogonadotropic hypogonadism. *The Journal of clinical endocrinology and metabolism* **100**, E646-654, doi:10.1210/jc.2015-1080 (2015).
- 131 Palmert, M. R. & Boepple, P. A. Variation in the timing of puberty: clinical spectrum and genetic investigation. *The Journal of clinical endocrinology and metabolism* **86**, 2364-2368, doi:10.1210/jcem.86.6.7603 (2001).
- 132 Waldstreicher, J. *et al.* The genetic and clinical heterogeneity of gonadotropin-releasing hormone deficiency in the human. *The Journal of clinical endocrinology and metabolism* **81**, 4388-4395, doi:10.1210/jcem.81.12.8954047 (1996).

- 133 Dwyer, A. A., Raivio, T. & Pitteloud, N. MANAGEMENT OF ENDOCRINE DISEASE: Reversible hypogonadotropic hypogonadism. *Eur J Endocrinol* **174**, R267-274, doi:10.1530/EJE-15-1033 (2016).
- 134 Harrington, J. & Palmert, M. R. Clinical review: Distinguishing constitutional delay of growth and puberty from isolated hypogonadotropic hypogonadism: critical appraisal of available diagnostic tests. *The Journal of clinical endocrinology and metabolism* **97**, 3056-3067, doi:10.1210/jc.2012-1598 (2012).
- 135 Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97, doi:10.1038/nature13545 (2014).
- 136 Day, F. R. *et al.* Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun* **6**, 8842, doi:10.1038/ncomms9842 (2015).
- 137 Lunetta, K. L. *et al.* Rare coding variants and X-linked loci associated with age at menarche. *Nat Commun* **6**, 7756, doi:10.1038/ncomms8756 (2015).
- 138 Lin, L. *et al.* A homozygous R262Q mutation in the gonadotropin-releasing hormone receptor presenting as constitutional delay of growth and puberty with subsequent borderline oligospermia. *The Journal of clinical endocrinology and metabolism* **91**, 5117-5121, doi:10.1210/jc.2006-0807 (2006).
- 139 Vaaralahti, K. *et al.* The role of gene defects underlying isolated hypogonadotropic hypogonadism in patients with constitutional delay of growth and puberty. *Fertil Steril* **95**, 2756-2758, doi:10.1016/j.fertnstert.2010.12.059 (2011).
- 140 Tusset, C. *et al.* Mutational analysis of TAC3 and TACR3 genes in patients with idiopathic central pubertal disorders. *Arq Bras Endocrinol Metabol* **56**, 646-652 (2012).
- 141 Howard, S. R. *et al.* IGSF10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. *EMBO Mol Med* **8**, 626-642, doi:10.15252/emmm.201606250 (2016).
- 142 Mitchell, A. L., Dwyer, A., Pitteloud, N. & Quinton, R. Genetic basis and variable phenotypic expression of Kallmann syndrome: towards a unifying theory. *Trends Endocrinol Metab* **22**, 249-258, doi:10.1016/j.tem.2011.03.002 (2011).
- 143 Wehkalampi, K., Widen, E., Laine, T., Palotie, A. & Dunkel, L. Association of the timing of puberty with a chromosome 2 locus. *The Journal of clinical endocrinology and metabolism* **93**, 4833-4839, doi:10.1210/jc.2008-0882 (2008).
- 144 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**, 377-394, doi:10.1089/1066527041410418 (2004).
- 145 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814 (2003).
- 146 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).
- 147 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190, doi:10.1038/nature12975 (2014).
- 148 Choi, J. H. *et al.* Expanding the Spectrum of Founder Mutations Causing Isolated Gonadotropin-Releasing Hormone Deficiency. *The Journal of clinical endocrinology and metabolism* **100**, E1378-1385, doi:10.1210/jc.2015-2262 (2015).
- 149 Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**, 198-199 (1998).

- 150 Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-511, doi:10.1038/nature12531 (2013).
- 151 Monnier, C. *et al.* PROKR2 missense mutations associated with Kallmann syndrome impair receptor signalling activity. *Hum Mol Genet* **18**, 75-81, doi:10.1093/hmg/ddn318 (2009).
- 152 Francou, B. *et al.* Normosmic congenital hypogonadotropic hypogonadism due to TAC3/TACR3 mutations: characterization of neuroendocrine phenotypes and novel mutations. *PLoS One* **6**, e25614, doi:10.1371/journal.pone.0025614 (2011).
- 153 Xu, N. *et al.* Nasal embryonic LHRH factor (NELF) mutations in patients with normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Fertil Steril* **95**, 1613-1620 e1611-1617, doi:10.1016/j.fertnstert.2011.01.010 (2011).
- 154 Amstalden, M. *et al.* Neurokinin 3 receptor immunoreactivity in the septal region, preoptic area and hypothalamus of the female sheep: colocalisation in neurokinin B cells of the arcuate nucleus but not in gonadotrophin-releasing hormone neurones. *J Neuroendocrinol* **22**, 1-12, doi:10.1111/j.1365-2826.2009.01930.x (2010).
- 155 Chan, Y. M. *et al.* GnRH-deficient phenotypes in humans and mice with heterozygous variants in KISS1/Kiss1. *The Journal of clinical endocrinology and metabolism* **96**, E1771-1781, doi:10.1210/jc.2011-0518 (2011).
- 156 Sanlaville, D. *et al.* Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human development. *J Med Genet* **43**, 211-217, doi:10.1136/jmg.2005.036160 (2006).
- 157 Gill, J. C., Moenter, S. M. & Tsai, P. S. Developmental regulation of gonadotropin-releasing hormone neurons by fibroblast growth factor signaling. *Endocrinology* **145**, 3830-3839, doi:10.1210/en.2004-0214 (2004).
- 158 Hardelin, J. P. & Dode, C. The complex genetics of Kallmann syndrome: KAL1, FGFR1, FGF8, PROKR2, PROK2, et al. *Sex Dev* **2**, 181-193, doi:10.1159/000152034 (2008).
- 159 de Castro, F. *et al.* The adhesion molecule anosmin-1 in neurology: Kallmann syndrome and beyond. *Adv Neurobiol* **8**, 273-292 (2014).
- 160 Legouis, R. *et al.* The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* **67**, 423-435 (1991).
- 161 Weller, S. & Gartner, J. Genetic and clinical aspects of X-linked hydrocephalus (L1 disease): Mutations in the L1CAM gene. *Hum Mutat* **18**, 1-12, doi:10.1002/humu.1144 (2001).
- 162 Gamble, J. A. *et al.* Disruption of ephrin signaling associates with disordered axophilic migration of the gonadotropin-releasing hormone neurons. *J Neurosci* **25**, 3142-3150, doi:10.1523/JNEUROSCI.4759-04.2005 (2005).
- 163 Yoshida, K., Rutishauser, U., Crandall, J. E. & Schwarting, G. A. Polysialic acid facilitates migration of luteinizing hormone-releasing hormone neurons on vomeronasal axons. *J Neurosci* **19**, 794-801 (1999).
- 164 Hrabovszky, E. *et al.* Sexual dimorphism of kisspeptin and neurokinin B immunoreactive neurons in the infundibular nucleus of aged men and women. *Front Endocrinol (Lausanne)* **2**, 80, doi:10.3389/fendo.2011.00080 (2011).
- 165 Law, C. *et al.* Normal Molecular Specification and Neurodegenerative Disease-Like Death of Spinal Neurons Lacking the SNARE-Associated Synaptic Protein Munc18-1. *J Neurosci* **36**, 561-576, doi:10.1523/JNEUROSCI.1964-15.2016 (2016).

- 166 Ramkhalawon, B. *et al.* Netrin-1 promotes adipose tissue macrophage retention and insulin resistance in obesity. *Nat Med* **20**, 377-384, doi:10.1038/nm.3467 (2014).
- 167 Bouilly, J. *et al.* New NOBOX mutations identified in a large cohort of women with primary ovarian insufficiency decrease KIT-L expression. *The Journal of clinical endocrinology and metabolism* **100**, 994-1001, doi:10.1210/jc.2014-2761 (2015).
- 168 Srour, M. *et al.* Mutations in DCC cause congenital mirror movements. *Science* **328**, 592, doi:10.1126/science.1186463 (2010).
- 169 Cirulli, V. & Yebra, M. Netrins: beyond the brain. *Nat Rev Mol Cell Biol* **8**, 296-306, doi:10.1038/nrm2142 (2007).
- 170 Schwarting, G. A., Kostek, C., Bless, E. P., Ahmad, N. & Tobet, S. A. Deleted in colorectal cancer (DCC) regulates the migration of luteinizing hormone-releasing hormone neurons to the basal forebrain. *J Neurosci* **21**, 911-919 (2001).
- 171 Schwarting, G. A., Raitcheva, D., Bless, E. P., Ackerman, S. L. & Tobet, S. Netrin 1-mediated chemoattraction regulates the migratory pathway of LHRH neurons. *Eur J Neurosci* **19**, 11-20 (2004).
- 172 Geisbrecht, B. V., Dowd, K. A., Barfield, R. W., Longo, P. A. & Leahy, D. J. Netrin binds discrete subdomains of DCC and UNC5 and mediates interactions between DCC and heparin. *J Biol Chem* **278**, 32561-32568, doi:10.1074/jbc.M302943200 (2003).
- 173 Cole, L. W. *et al.* Mutations in prokineticin 2 and prokineticin receptor 2 genes in human gonadotrophin-releasing hormone deficiency: molecular genetics and clinical spectrum. *The Journal of clinical endocrinology and metabolism* **93**, 3551-3559, doi:10.1210/jc.2007-2654 (2008).
- 174 Sidhoum, V. F. *et al.* Reversal and relapse of hypogonadotropic hypogonadism: resilience and fragility of the reproductive neuroendocrine system. *The Journal of clinical endocrinology and metabolism* **99**, 861-870, doi:10.1210/jc.2013-2809 (2014).
- 175 Casoni, F. *et al.* Development of the neurons controlling fertility in humans: new insights from 3D imaging and transparent fetal brains. *Development* **143**, 3969-3981, doi:10.1242/dev.139444 (2016).
- 176 Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61-85 (1990).
- 177 Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175-185 (1996).
- 178 Lakhina, V. *et al.* Netrin/DCC signaling guides olfactory sensory axons to their correct location in the olfactory bulb. *J Neurosci* **32**, 4440-4456, doi:10.1523/JNEUROSCI.4442-11.2012 (2012).
- 179 Depienne, C. *et al.* A novel DCC mutation and genetic heterogeneity in congenital mirror movements. *Neurology* **76**, 260-264, doi:10.1212/WNL.0b013e318207b1e0 (2011).
- 180 Meneret, A. *et al.* Congenital mirror movements: mutational analysis of RAD51 and DCC in 26 cases. *Neurology* **82**, 1999-2002, doi:10.1212/WNL.0000000000000477 (2014).
- 181 Marsh, A. P. *et al.* Mutations in DCC cause isolated agenesis of the corpus callosum with incomplete penetrance. *Nat Genet* **49**, 511-514, doi:10.1038/ng.3794 (2017).

- 182 Jamuar, S. S. *et al.* Biallelic mutations in human DCC cause developmental split-brain syndrome. *Nat Genet* **49**, 606-612, doi:10.1038/ng.3804 (2017).
- 183 Chen, Q. *et al.* N-terminal horseshoe conformation of DCC is functionally required for axon guidance and might be shared by other neural receptors. *J Cell Sci* **126**, 186-195, doi:10.1242/jcs.111278 (2013).
- 184 Xu, K. *et al.* Neural migration. Structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism. *Science* **344**, 1275-1279, doi:10.1126/science.1255149 (2014).
- 185 Finci, L., Zhang, Y., Meijers, R. & Wang, J. H. Signaling mechanism of the netrin-1 receptor DCC in axon guidance. *Prog Biophys Mol Biol* **118**, 153-160, doi:10.1016/j.pbiomolbio.2015.04.001 (2015).
- 186 Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nature reviews. Genetics* **13**, 565-575, doi:10.1038/nrg3241 (2012).
- 187 Acuna-Hidalgo, R., Veltman, J. A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol* **17**, 241, doi:10.1186/s13059-016-1110-1 (2016).
- 188 Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature reviews. Genetics* **8**, 610-618, doi:10.1038/nrg2146 (2007).
- 189 Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nature reviews. Genetics* **1**, 40-47, doi:10.1038/35049558 (2000).
- 190 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445-449, doi:10.1126/science.1138659 (2007).
- 191 Xu, B. *et al.* Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* **40**, 880-885, doi:10.1038/ng.162 (2008).
- 192 Hoischen, A. *et al.* De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* **42**, 483-485, doi:10.1038/ng.581 (2010).
- 193 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793, doi:10.1038/ng.646 (2010).
- 194 Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-241, doi:10.1038/nature10945 (2012).
- 195 Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-245, doi:10.1038/nature11011 (2012).
- 196 O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-250, doi:10.1038/nature10989 (2012).
- 197 Hu, H. *et al.* X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol Psychiatry* **21**, 133-148, doi:10.1038/mp.2014.193 (2016).
- 198 Simpson, M. A. *et al.* Mutations in NOTCH2 cause Hajdu-Cheney syndrome, a disorder of severe and progressive bone loss. *Nat Genet* **43**, 303-305, doi:10.1038/ng.779 (2011).
- 199 Vissers, L. E. *et al.* Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, encoding the Golgi-resident nucleotide phosphatase, gPAPP. *Am J Hum Genet* **88**, 608-615, doi:10.1016/j.ajhg.2011.04.002 (2011).

- 200 Auer, P. L. & Lettre, G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* **7**, 16, doi:10.1186/s13073-015-0138-2 (2015).
- 201 Zhu, N. *et al.* Strategies to improve the performance of rare variant association studies by optimizing the selection of controls. *Bioinformatics* **31**, 3577-3583, doi:10.1093/bioinformatics/btv457 (2015).
- 202 Wei, Q. *et al.* A Bayesian framework for de novo mutation calling in parents-offspring trios. *Bioinformatics*, doi:10.1093/bioinformatics/btu839 (2014).
- 203 Fuentes Fajardo, K. V. *et al.* Detecting false-positive signals in exome sequencing. *Hum Mutat* **33**, 609-613, doi:10.1002/humu.22033 (2012).
- 204 Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003-1007, doi:10.1126/science.1072047 (2002).
- 205 Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome research* **11**, 1005-1017, doi:10.1101/gr.187101 (2001).
- 206 Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873, doi:10.1093/bioinformatics/btq559 (2010).
- 207 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).
- 208 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
- 209 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
- 210 Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* **100**, 267-280, doi:10.1016/j.ajhg.2017.01.004 (2017).
- 211 Messina, A. *et al.* A microRNA switch regulates the rise in hypothalamic GnRH production before puberty. *Nat Neurosci* **19**, 835-844, doi:10.1038/nn.4298 (2016).
- 212 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).
- 213 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).
- 214 Turner, S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*, doi:10.1101/005165 (2014).
- 215 Cariboni, A. *et al.* Slit2 and Robo3 modulate the migration of GnRH-secreting neurons. *Development* **139**, 3326-3331, doi:10.1242/dev.079418 (2012).
- 216 Tetreault, M. *et al.* Recessive mutations in POLR3B, encoding the second largest subunit of Pol III, cause a rare hypomyelinating leukodystrophy. *Am J Hum Genet* **89**, 652-655, doi:10.1016/j.ajhg.2011.10.006 (2011).

- 217 Saitsu, H. *et al.* Mutations in POLR3A and POLR3B encoding RNA Polymerase III subunits cause an autosomal-recessive hypomyelinating leukoencephalopathy. *Am J Hum Genet* **89**, 644-651, doi:10.1016/j.ajhg.2011.10.003 (2011).
- 218 Richards, M. R. *et al.* Phenotypic spectrum of POLR3B mutations: isolated hypogonadotropic hypogonadism without neurological or dental anomalies. *J Med Genet* **54**, 19-25, doi:10.1136/jmedgenet-2016-104064 (2017).
- 219 Yip, B., Chen, S. H., Mulder, H., Hoppener, J. W. & Schachter, H. Organization of the human beta-1,2-N-acetylglucosaminyltransferase I gene (MGAT1), which controls complex and hybrid N-glycan synthesis. *Biochem J* **321 (Pt 2)**, 465-474 (1997).
- 220 Ioffe, E. & Stanley, P. Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc Natl Acad Sci U S A* **91**, 728-732 (1994).
- 221 Williams, S. A. & Stanley, P. Oocyte-specific deletion of complex and hybrid N-glycans leads to defects in preovulatory follicle and cumulus mass development. *Reproduction* **137**, 321-331, doi:10.1530/REP-07-0469 (2009).
- 222 Ye, Z. & Marth, J. D. N-glycan branching requirement in neuronal and postnatal viability. *Glycobiology* **14**, 547-558, doi:10.1093/glycob/cwh069 (2004).
- 223 Gil-Rodriguez, M. C. *et al.* De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. *Hum Mutat* **36**, 454-462, doi:10.1002/humu.22761 (2015).
- 224 Losada, A. Cohesin in cancer: chromosome segregation and beyond. *Nat Rev Cancer* **14**, 389-393, doi:10.1038/nrc3743 (2014).
- 225 Deardorff, M. A. *et al.* Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet* **80**, 485-494, doi:10.1086/511888 (2007).
- 226 Kline, A. D. *et al.* Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. *Am J Med Genet A* **143A**, 1287-1296, doi:10.1002/ajmg.a.31757 (2007).
- 227 Mariani, M. *et al.* Adolescents and adults affected by Cornelia de Lange syndrome: A report of 73 Italian patients. *Am J Med Genet C Semin Med Genet* **172**, 206-213, doi:10.1002/ajmg.c.31502 (2016).
- 228 Oliver, C. *et al.* Cornelia de Lange syndrome: extending the physical and psychological phenotype. *Am J Med Genet A* **152A**, 1127-1135, doi:10.1002/ajmg.a.33363 (2010).
- 229 Kline, A. D. *et al.* Developmental data on individuals with the Brachmann-de Lange syndrome. *Am J Med Genet* **47**, 1053-1058, doi:10.1002/ajmg.1320470724 (1993).
- 230 Liu, J. & Krantz, I. D. Cornelia de Lange syndrome, cohesin, and beyond. *Clin Genet* **76**, 303-314, doi:10.1111/j.1399-0004.2009.01271.x (2009).
- 231 Ramos, F. J. *et al.* Clinical utility gene card for: Cornelia de Lange syndrome. *European journal of human genetics : EJHG* **23**, doi:10.1038/ejhg.2014.270 (2015).
- 232 Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature reviews. Genetics* **14**, 681-691, doi:10.1038/nrg3555 (2013).

- 233 Abreu, A. P. *et al.* Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med* **368**, 2467-2475, doi:10.1056/NEJMoa1302160 (2013).
- 234 Cariboni, A. *et al.* Dysfunctional SEMA3E signaling underlies gonadotropin-releasing hormone neuron deficiency in Kallmann syndrome. *J Clin Invest* **125**, 2413-2428, doi:10.1172/JCI78448 (2015).
- 235 Niclou, S. P., Jia, L. & Raper, J. A. Slit2 is a repellent for retinal ganglion cell axons. *J Neurosci* **20**, 4962-4974 (2000).
- 236 Xu, H. T. *et al.* Calcium signaling in chemorepellant Slit2-dependent regulation of neuronal migration. *Proc Natl Acad Sci U S A* **101**, 4296-4301, doi:10.1073/pnas.0303893101 (2004).
- 237 Fothergill, T. *et al.* Netrin-DCC signaling regulates corpus callosum formation through attraction of pioneering axons and by modulating Slit2-mediated repulsion. *Cereb Cortex* **24**, 1138-1151, doi:10.1093/cercor/bhs395 (2014).
- 238 Gibson, D. A. *et al.* Dendrite self-avoidance requires cell-autonomous slit/robo signaling in cerebellar purkinje cells. *Neuron* **81**, 1040-1056, doi:10.1016/j.neuron.2014.01.009 (2014).
- 239 Nguyen-Ba-Charvet, K. T., Plump, A. S., Tessier-Lavigne, M. & Chedotal, A. Slit1 and slit2 proteins control the development of the lateral olfactory tract. *J Neurosci* **22**, 5473-5480 (2002).
- 240 Svensson, K. J. *et al.* A Secreted Slit2 Fragment Regulates Adipose Tissue Thermogenesis and Metabolic Function. *Cell Metab* **23**, 454-466, doi:10.1016/j.cmet.2016.01.008 (2016).
- 241 Bousfield, G. R. & Dias, J. A. Synthesis and secretion of gonadotropins including structure-function correlates. *Rev Endocr Metab Disord* **12**, 289-302, doi:10.1007/s11154-011-9191-3 (2011).
- 242 Metzler, M. *et al.* Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. *EMBO J* **13**, 2056-2065 (1994).
- 243 Shi, S. *et al.* Inactivation of the Mgat1 gene in oocytes impairs oogenesis, but embryos lacking complex and hybrid N-glycans develop and implant. *Mol Cell Biol* **24**, 9920-9929, doi:10.1128/MCB.24.22.9920-9929.2004 (2004).
- 244 White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452-464, doi:10.1016/j.cell.2013.06.022 (2013).
- 245 Acampora, D. *et al.* Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5. *Development* **126**, 3795-3809 (1999).
- 246 Nishiyama, T., Sykora, M. M., Huis in 't Veld, P. J., Mechtler, K. & Peters, J. M. Aurora B and Cdk1 mediate Wapl activation and release of acetylated cohesin from chromosomes by phosphorylating Sororin. *Proc Natl Acad Sci U S A* **110**, 13404-13409, doi:10.1073/pnas.1305020110 (2013).
- 247 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 248 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).

- 249 Yuan, B. *et al.* Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *J Clin Invest* **125**, 636-651, doi:10.1172/JCI77435 (2015).
- 250 Rohatgi, S. *et al.* Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey. *Am J Med Genet A* **152A**, 1641-1653, doi:10.1002/ajmg.a.33441 (2010).
- 251 Selicorni, A. *et al.* Clinical score of 62 Italian patients with Cornelia de Lange syndrome and correlations with the presence and type of NIPBL mutation. *Clin Genet* **72**, 98-108, doi:10.1111/j.1399-0004.2007.00832.x (2007).
- 252 Ghiselli, G. & Liu, C. G. Global gene expression profiling of cells overexpressing SMC3. *Mol Cancer* **4**, 34, doi:10.1186/1476-4598-4-34 (2005).
- 253 Choi, S. G. *et al.* Growth differentiation factor 9 (GDF9) forms an incoherent feed-forward loop modulating follicle-stimulating hormone beta-subunit (FSHbeta) gene expression. *J Biol Chem* **289**, 16164-16175, doi:10.1074/jbc.M113.537696 (2014).
- 254 Pistocchi, A. *et al.* Cornelia de Lange Syndrome: NIPBL haploinsufficiency downregulates canonical Wnt pathway in zebrafish embryos and patients fibroblasts. *Cell Death Dis* **4**, e866, doi:10.1038/cddis.2013.371 (2013).
- 255 Zaghetto, A. A. *et al.* Activation of the Wnt-beta catenin pathway in a cell population on the surface of the forebrain is essential for the establishment of olfactory axon connections. *J Neurosci* **27**, 9757-9768, doi:10.1523/JNEUROSCI.0763-07.2007 (2007).
- 256 Aujla, P. K., Bora, A., Monahan, P., Sweedler, J. V. & Raetzman, L. T. The Notch effector gene Hes1 regulates migration of hypothalamic neurons, neuropeptide content and axon targeting to the pituitary. *Dev Biol* **353**, 61-71, doi:10.1016/j.ydbio.2011.02.018 (2011).
- 257 Lund, C. *et al.* Development of Gonadotropin-Releasing Hormone-Secreting Neurons from Human Pluripotent Stem Cells. *Stem Cell Reports* **7**, 149-157, doi:10.1016/j.stemcr.2016.06.007 (2016).
- 258 Basu, S. & Pan, W. Comparison of statistical tests for disease association with rare variants. *Genet Epidemiol* **35**, 606-619, doi:10.1002/gepi.20609 (2011).
- 259 Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762-775, doi:10.1093/biostatistics/kxs014 (2012).
- 260 Goldstein, D. B. *et al.* Sequencing studies in human genetics: design and interpretation. *Nature reviews. Genetics* **14**, 460-470, doi:10.1038/nrg3455 (2013).
- 261 Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93, doi:10.1016/j.ajhg.2011.05.029 (2011).
- 262 Neale, B. M. *et al.* Testing for an unusual distribution of rare variants. *PLoS Genet* **7**, e1001322, doi:10.1371/journal.pgen.1001322 (2011).
- 263 Pan, W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol* **33**, 497-507, doi:10.1002/gepi.20402 (2009).
- 264 Santorico, S. A. & Hendricks, A. E. Progress in methods for rare variant association. *BMC Genet* **17 Suppl 2**, 6, doi:10.1186/s12863-015-0316-7 (2016).
- 265 Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23, doi:10.1016/j.ajhg.2014.06.009 (2014).

- 266 Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**, 224-237, doi:10.1016/j.ajhg.2012.06.007 (2012).
- 267 Sun, J., Zheng, Y. & Hsu, L. A unified mixed-effects model for rare-variant association in sequencing studies. *Genet Epidemiol* **37**, 334-344, doi:10.1002/gepi.21717 (2013).
- 268 Samuels, M. E. Saturation of the human phenome. *Curr Genomics* **11**, 482-499, doi:10.2174/138920210793175886 (2010).
- 269 Chan, P. A. *et al.* Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat* **28**, 683-693, doi:10.1002/humu.20492 (2007).
- 270 Hicks, S., Wheeler, D. A., Plon, S. E. & Kimmel, M. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat* **32**, 661-668, doi:10.1002/humu.21490 (2011).
- 271 Vissers, L. E., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nature reviews. Genetics* **17**, 9-18, doi:10.1038/nrg3999 (2016).
- 272 Loviglio, M. N. *et al.* Identification of a RAI1-associated disease network through integration of exome sequencing, transcriptomics, and 3D genomics. *Genome Med* **8**, 105, doi:10.1186/s13073-016-0359-z (2016).
- 273 Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81, doi:10.1038/nature15394 (2015).
- 274 Stephens, Z. D. *et al.* Big Data: Astronomical or Genomical? *PLoS Biol* **13**, e1002195, doi:10.1371/journal.pbio.1002195 (2015).
- 275 Linder, J. A. & Levine, D. M. Health Care Communication Technology and Improved Access, Continuity, and Relationships The Revolution Will Be Uberized. *Jama Intern Med* **176**, 643-644, doi:10.1001/jamainternmed.2016.0692 (2016).
- 276 Democratizing sequencing. *Nature methods* **2**, 633, doi:10.1038/nmeth0905-633 (2005).
- 277 Hood, L. & Price, N. D. Demystifying disease, democratizing health care. *Sci Transl Med* **6**, 225ed225, doi:10.1126/scitranslmed.3008665 (2014).
- 278 Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* **46**, 61-64, doi:10.1038/ng.2826 (2014).
- 279 Duan, J. *et al.* A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *Am J Hum Genet* **95**, 744-753, doi:10.1016/j.ajhg.2014.11.001 (2014).
- 280 Bakker, J. L. *et al.* A novel splice site mutation in the noncoding region of BRCA2: implications for Fanconi anemia and familial breast cancer diagnostics. *Hum Mutat* **35**, 442-446, doi:10.1002/humu.22505 (2014).
- 281 Scacheri, C. A. & Scacheri, P. C. Mutations in the noncoding genome. *Curr Opin Pediatr* **27**, 659-664, doi:10.1097/MOP.0000000000000283 (2015).

Supplementary Material

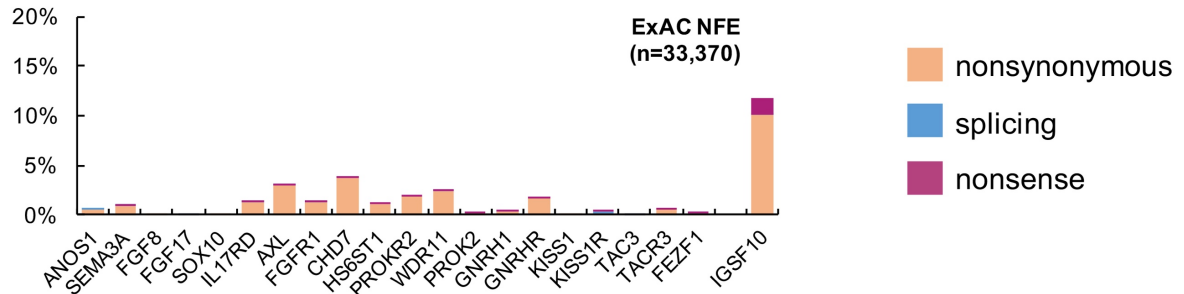


Figure S1. ExAC European individuals show marginal mutation prevalence in CHH genes. Histograms showing CHH genes mutational estimated prevalence in ExAC non-Finnish European (n=33,370). Each bar contains the frequency of nonsynonymous (orange), splicing (blue) and nonsense (purple) variants accounting for each gene prevalence. Estimation of prevalence was calculated from the number of heterozygous and homozygous variants found in each gene.

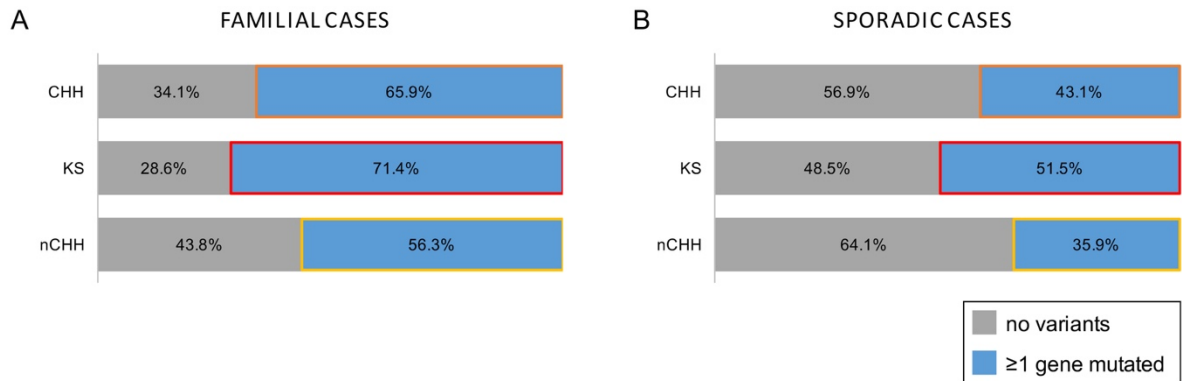


Figure S2. Familial and sporadic cases display different burden of rare variants in CHH genes. Frequencies of familial (A) and sporadic (B) CHH, KS and nCHH probands with no mutations (grey) and with at least one mutation in CHH genes.

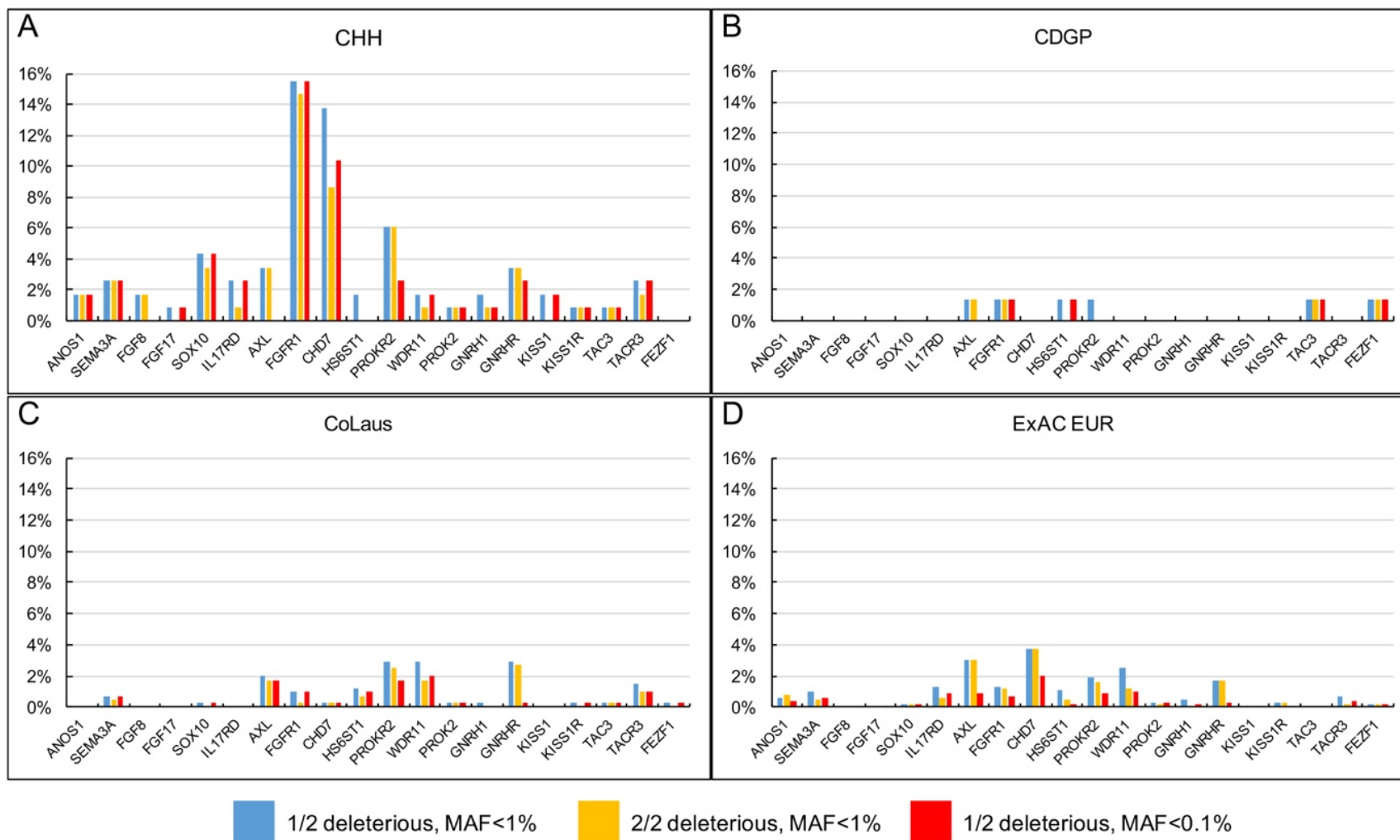


Figure S3. Different strategies in variants filtering. Mutation prevalence of known CHH genes in (A) CHH, (B) CDGP, (C) CoLaus and (D) ExAC NFE individuals using three different filtering strategies: (1) MAF < 1% in ExAC NFE and at least one deleterious prediction in SIFT and/or PolyPhen-2 for missense variants (blue bars); (2) MAF < 1% in ExAC NFE and two deleterious predictions in SIFT and PolyPhen-2 for missense variants (yellow bars); (3) MAF < 0.1% in ExAC NFE and at least one deleterious prediction in SIFT and/or PolyPhen-2 for missense variants (red bars).

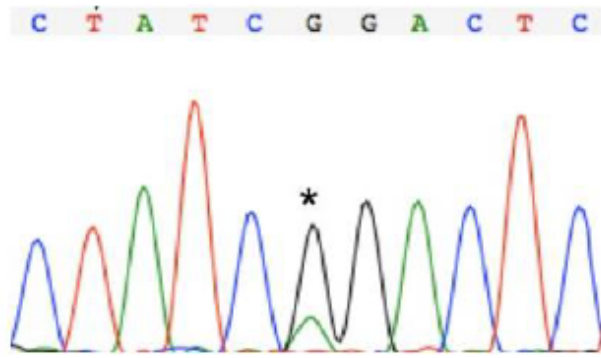


Figure S4. Sanger sequencing chromatogram showing *FGFR1 de novo*, putative mosaic variant.

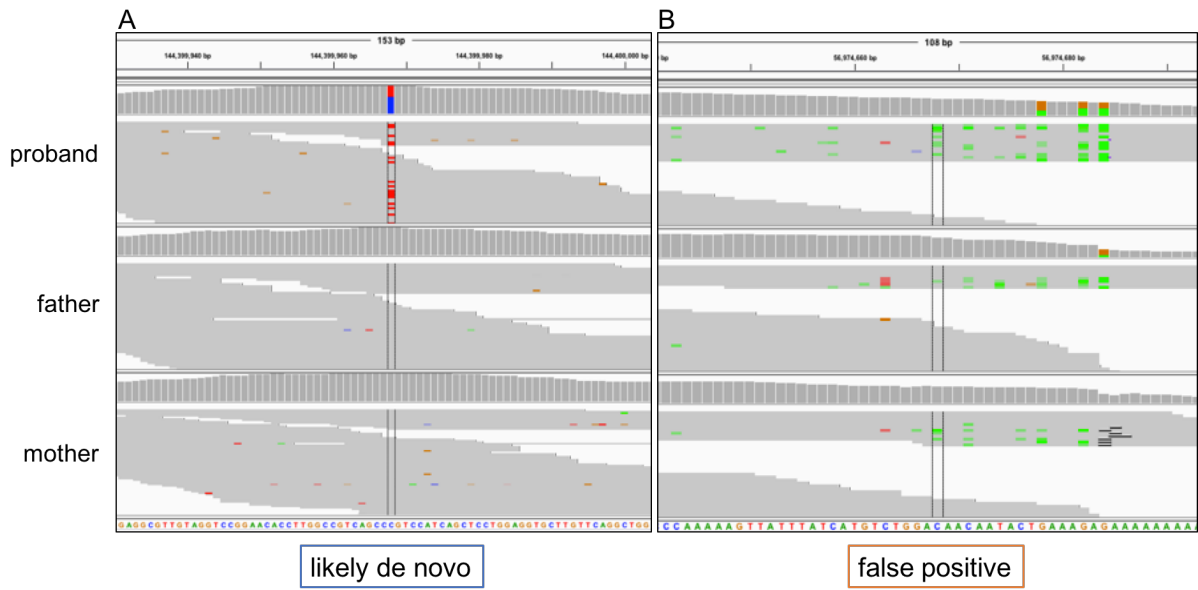


Figure S5. Examples of a true positive and false positive de novo calls by TrioDeNovo on IGV.

Table S1. OMIM entries with “hypogonadotropic hypogonadism” in clinical synopses.

MIM Number	Title
#101800	ACRODYSOSTOSIS 1 WITH OR WITHOUT HORMONE RESISTANCE; ACRDYS1
#609441	ACROMESOMELIC DYSPLASIA, DEMIRHAN TYPE; AMDD
#612079	ALOPECIA, NEUROLOGIC DEFECTS, AND ENDOCRINOPATHY SYNDROME; ANES
#614307	ALPHA-METHYLACYL-CoA RACEMASE DEFICIENCY; AMACRD
#203800	ALSTROM SYNDROME; ALMS
#312300	ANDROGEN INSENSITIVITY, PARTIAL; PAIS
#615234	ANEMIA, HYPOCHROMIC MICROCYTIC, WITH IRON OVERLOAD 2; AHMIO2
#208900	ATAXIA-TELANGIECTASIA; AT
#240300	AUTOIMMUNE POLYENDOCRINE SYNDROME, TYPE I, WITH OR WITHOUT REVERSIBLE METAPHYSEAL DYSPLASIA; APS1
#209900	BARDET-BIEDL SYNDROME 1; BBS1
#615987	BARDET-BIEDL SYNDROME 10; BBS10
#615988	BARDET-BIEDL SYNDROME 11; BBS11
#615989	BARDET-BIEDL SYNDROME 12; BBS12
#615993	BARDET-BIEDL SYNDROME 16; BBS16
#615994	BARDET-BIEDL SYNDROME 17; BBS17
#615996	BARDET-BIEDL SYNDROME 19; BBS19
#615981	BARDET-BIEDL SYNDROME 2; BBS2
#615982	BARDET-BIEDL SYNDROME 4; BBS4
#615983	BARDET-BIEDL SYNDROME 5; BBS5
#615985	BARDET-BIEDL SYNDROME 8; BBS8
#215470	BOUCHER-NEUHAUSER SYNDROME; BNHS
#212112	CARDIOMYOPATHY, DILATED, WITH HYPERGONADOTROPIC HYPOGONADISM
#214800	CHARGE SYNDROME
#302950	CHONDRODYSPLASIA PUNCTATA 1, X-LINKED RECESSIVE; CDPX1
#612513	CHROMOSOME 2p16.1-p15 DELETION SYNDROME
#300869	CHROMOSOME Xq27.3-q28 DUPLICATION SYNDROME
#216400	COCKAYNE SYNDROME A; CSA
#604168	CONGENITAL CATARACTS, FACIAL DYSMORPHISM, AND NEUROPATHY; CCFDN
#212065	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ia; CDG1A
#608540	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE Ik; CDG1K
#614921	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE It; CDG1T
#300882	CORNELIA DE LANGE SYNDROME 5; CDLS5
#615849	CULLER-JONES SYNDROME; CJS
#604292	ECTRODACTYLY, ECTODERMAL DYSPLASIA, AND CLEFT LIP/PALATE SYNDROME 3; EEC3
#227650	FANCONI ANEMIA, COMPLEMENTATION GROUP A; FANCA
#227646	FANCONI ANEMIA, COMPLEMENTATION GROUP D2; FANCD2
#600901	FANCONI ANEMIA, COMPLEMENTATION GROUP E; FANCE
#230400	GALACTOSEMIA
#212840	GORDON HOLMES SYNDROME; GDHS
#615465	HARTSFIELD SYNDROME; HRTFDS
#235200	HEMOCHROMATOSIS, TYPE 1; HFE1
#602782	HISTIOCYTOSIS-LYMPHADENOPATHY PLUS SYNDROME

#606407	HYPOTONIA-CYSTINURIA SYNDROME
#308100	ICHTHYOSIS, X-LINKED; XLI
#614962	LEPTIN DEFICIENCY OR DYSFUNCTION; LEPD
#607694	LEUKODYSTROPHY, HYPOMYELINATING, 7, WITH OR WITHOUT OLIGODONTIA AND/OR HYPOGONADOTROPIC HYPOGONADISM; HLD7
#614381	LEUKODYSTROPHY, HYPOMYELINATING, 8, WITH OR WITHOUT OLIGODONTIA AND/OR HYPOGONADOTROPIC HYPOGONADISM; HLD8
#613724	LEUKOENCEPHALOPATHY WITH DYSTONIA AND MOTOR NEUROPATHY
#613075	MACS SYNDROME
#615381	MANDIBULAR HYPOPLASIA, DEAFNESS, PROGEROID FEATURES, AND LIPODYSTROPHY SYNDROME; MDPL
#248800	MARINESCO-SJOGREN SYNDROME; MSS
#212720	MARTSOLF SYNDROME
#309580	MENTAL RETARDATION-HYPOTONIC FACIES SYNDROME, X-LINKED, 1; MRXHF1
#300987	MENTAL RETARDATION, X-LINKED, SYNDROMIC, BORCK TYPE; MRXSBRK
#300354	MENTAL RETARDATION, X-LINKED, SYNDROMIC, CABEZAS TYPE; MRXSC
#614231	MICROCEPHALY, EPILEPSY, AND DIABETES SYNDROME; MEDS
#206900	MICROPTHALMIA, SYNDROMIC 3; MCOPS3
#617053	MIRAGE SYNDROME; MIRAGE
#615084	MITOCHONDRIAL DNA DEPLETION SYNDROME 11; MTDPS11
#271245	MITOCHONDRIAL DNA DEPLETION SYNDROME 7 (HEPATOCEREBRAL TYPE); MTDPS7
#300845	MOYAMOYA DISEASE 4 WITH SHORT STATURE, HYPERGONADOTROPIC HYPOGONADISM, AND FACIAL DYSMORPHISM; MYMY4
#160900	MYOTONIC DYSTROPHY 1; DM1
#602668	MYOTONIC DYSTROPHY 2; DM2
#163950	NOONAN SYNDROME 1; NS1
#275400	OLIVER-MCFARLANE SYNDROME; OMCS
#616138	PERRAULT SYNDROME 5; PRLTS5
#262600	PITUITARY HORMONE DEFICIENCY, COMBINED, 2; CPHD2
#182230	PITUITARY HORMONE DEFICIENCY, COMBINED, 5, INCLUDED; CPHD5
#616113	POLYENDOCRINE-POLYNEUROPATHY SYNDROME; PEPNS
#176270	PRADER-WILLI SYNDROME; PWS
#259050	PRIMROSE SYNDROME; PRIMS
#157640	PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA WITH MITOCHONDRIAL DNA DELETIONS, AUTOSOMAL DOMINANT 1; PEOA1
#609286	PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA WITH MITOCHONDRIAL DNA DELETIONS, AUTOSOMAL DOMINANT 3; PEOA3
#600955	PROPROTEIN CONVERTASE 1/3 DEFICIENCY
#103580	PSEUDOHYPOPARATHYROIDISM, TYPE IA; PHP1A
#612462	PSEUDOHYPOPARATHYROIDISM, TYPE IC; PHP1C
#268400	ROTHMUND-THOMSON SYNDROME; RTS
#615547	SCHAAF-YANG SYNDROME; SHFYNG
#616629	SENIOR-LOKEN SYNDROME 9; SLSN9
#617159	SIFRIM-HITZ-WEISS SYNDROME; SIHIWES
#615768	SPINOCEREBELLAR ATAXIA, AUTOSOMAL RECESSIVE 16; SCAR16
#613266	WAARDENBURG SYNDROME, TYPE 4C; WS4C
#277700	WERNER SYNDROME; WRN
#613406	WITTEVEEN-KOLK SYNDROME; WITKOS

#241080 WOODHOUSE-SAKATI SYNDROME

#610651 XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP B; XPB

Table S2. Mutation prevalence of CHH genes in screened cohorts.

Gene	CHH	KS	nCHH	CDGP	CoLaus	1000 Genomes EUR	ExAC NFE
<i>ANOS1</i>	1.7%	3.3%	0.0%	0.0%	0.0%	1.0%	0.6%
<i>SEMA3A</i>	2.6%	4.9%	0.0%	0.0%	0.7%	1.0%	1.0%
<i>FGF8</i>	1.7%	3.3%	0.0%	0.0%	0.0%	0.0%	0.1%
<i>FGF17</i>	0.9%	1.6%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>SOX10</i>	4.3%	6.6%	1.8%	0.0%	0.2%	0.0%	0.2%
<i>IL17RD</i>	2.6%	3.3%	1.8%	0.0%	0.0%	1.5%	1.3%
<i>AXL</i>	3.4%	1.6%	5.5%	1.4%	2.0%	2.0%	3.0%
<i>FGFR1</i>	15.5%	19.7%	10.9%	1.4%	1.0%	0.5%	1.3%
<i>CHD7</i>	13.8%	18.0%	9.1%	0.0%	0.2%	2.5%	3.7%
<i>HS6ST1</i>	1.7%	1.6%	1.8%	1.4%	1.2%	1.0%	1.1%
<i>PROKR2</i>	5.2%	8.2%	1.8%	1.4%	3.0%	2.0%	1.9%
<i>WDR11</i>	1.7%	1.6%	1.8%	0.0%	3.0%	1.5%	2.5%
<i>PROK2</i>	0.9%	0.0%	1.8%	0.0%	0.2%	0.0%	0.2%
<i>GNRH1</i>	1.7%	0.0%	3.6%	0.0%	0.2%	0.0%	0.5%
<i>GNRHR</i>	3.4%	0.0%	7.3%	0.0%	3.0%	0.5%	1.7%
<i>KISS1</i>	1.7%	0.0%	3.6%	0.0%	0.0%	0.0%	0.1%
<i>KISS1R</i>	0.9%	0.0%	1.8%	0.0%	0.2%	0.5%	0.3%
<i>TAC3</i>	0.9%	0.0%	1.8%	1.4%	0.2%	0.0%	0.1%
<i>TACR3</i>	2.6%	0.0%	5.5%	0.0%	1.5%	1.0%	0.7%
<i>FEZF1</i>	0.0%	0.0%	0.0%	1.4%	0.2%	0.5%	0.2%

Prevalence of putative mutations in cases and controls. ExAC NFE prevalence was estimated by dividing the sum of heterozygous and homozygous mutations in each gene to the total population (n=33,370).

Table S3. Putative mutations identified in the CHH cohort.

Sample	Phenotype	Gene	nt change	aa change	Zyg	SIFT	PPH2	MaxEnt	In vitro	ExAC	ExAC NFE	Previous report
1	KS	<i>CHD7</i>	c.3056T>G	p.Phe1019Cys	Het	D	D			private	private	
2	KS	<i>CHD7</i>	c.1397C>T	p.Ser466Leu	Het	D	T			0.11%	0.20%	Felix 2006, AJMG
3	KS	<i>CHD7</i>	c.4914T>G	p.Asp1638Glu	Het	D	D			0.00%	0.00%	
4	KS	<i>FGFR1</i>	c.2058delC	p.Phe686fs	Het					private	private	
5	KS	<i>FGFR1</i>	c.670G>C	p.Asp224His	Het	D	D			private	private	
6	KS	<i>SEMA3A</i>	c.2201G>A	p.Arg734Gln	Het	D	D			0.00%	0.00%	
7	nCHH	<i>CHD7</i>	c.7199G>A	p.Arg2400Gln	Het	T	D			0.00%	0.00%	
8	nCHH	<i>PROKR2</i>	c.1019C>G	p.Thr340Ser	Het	D	D			private	private	
			c.332T>G	p.Met111Arg	Het	D	D			private	private	
9	nCHH	<i>GNRHR</i>	c.785G>A	p.Arg262Gln	Hom	D	D			0.20%	0.24%	
		<i>CHD7</i>	c.8950C>T	p.Leu2984Phe	Het	D	T			0.45%	0.97%	
10	KS	<i>KAL1</i>	c.1756C>T	p.Gln586X	Hem					private	private	Miraoui 2013, AJHG
11	KS	<i>SOX10</i>	c.267delC	p.Met90fs	Het					private	private	
12	KS	<i>PROKR2</i>	c.518T>G	p.Leu173Arg	Het	D	D		LOF	0.22%	0.35%	Reynaud 2012, JCEM
		<i>SEMA3A</i>	c.196C>T	p.Arg66Trp	Het	D	D			0.05%	0.08%	Hanchate 2012, Plos Genet
13	KS	<i>CHD7</i>	c.4847A>G	p.Tyr1616Cys	Het	D	D			private	private	Balasubramanian 2014, PNAS
		<i>IL17RD</i>	c.1690T>G	p.Phe564Val	Het	T	D			0.01%	0.01%	
14	KS	<i>FGFR1</i>	c.232C>T	p.Arg78Cys	Het	D	D			private	private	Pitteloud 2006, Mol Cell Endocrinol

15	KS	<i>FGF17</i>	c.287G>A	p.Arg96Gln	Het	T	D		private	private	
16	KS	<i>FGFR1</i>	c.1430+1delG		Het			-207.8%	private	private	
		<i>CHD7</i>	c.1105C>G	p.Pro369Ala	Het	T	D		0.00%	0.00%	
17	KS	<i>PROKR2</i>	c.518T>G	p.Leu173Arg	Het	D	D	LOF	0.22%	0.35%	Reynaud 2012, JCEM
18	nCHH	<i>AXL</i>	c.1549G>A	p.Gly517Ser	Het	D	D		0.45%	0.68%	
19	nCHH	<i>GNRH1</i>	c.87delA	p.Leu30fs	Hom				private	private	
20	nCHH	<i>TACR3</i>	c.443A>T	p.His148Leu	Hom	D	D	LOF	private	private	Guran 2009, JCEM
21	nCHH	<i>FGFR1</i>	c.1756_1763dupAACCCAG	p.Ser588fs	Het				private	private	
22	KS	<i>SEMA3A</i>	c.1360+2T>G		Het			-87.2%	private	private	
23	KS	<i>IL17RD</i>	c.1136A>G	p.Tyr379Cys	Het	T	D	LOF	0.01%	0.01%	Miraoui 2013, AJHG
		<i>FGFR1</i>	c.1042G>A	p.Gly348Arg	Het	D	D		private	private	
24	KS	<i>WDR11</i>	c.1342C>T	p.Arg448Trp	Het	D	D		0.01%	0.00%	
25	nCHH	<i>GNRHR</i>	c.31C>A	p.Gln11Lys	Het	T	T	LOF	0.02%	0.03%	Meysing 2004, JCEM
			c.30T>A	p.Asn10Lys	Het	T	T	LOF	0.02%	0.03%	Gianetti 2010, JCEM
		<i>WDR11</i>	c.1342C>T	p.Arg448Trp	Het	D	D		0.01%	0.00%	
26	nCHH	<i>TACR3</i>	c.824G>A	p.Trp275X	Hom				0.02%	0.04%	Gianetti 2010, JCEM
27	nCHH	<i>KISS1R</i>	c.772C>T	p.Arg258Trp	Het	D	D		0.00%	0.00%	
		<i>SOX10</i>	c.481C>T	p.Arg161Cys	Het	D	D		private	private	

28	KS	<i>FGFR1</i>	c.1042G>A	p.Gly348Arg	Het	D	D		private	private	Bailleul-Forestier 2010, Int J Paediatr Dent
29	KS	<i>AXL</i>	c.1549G>A	p.Gly517Ser	Het	D	D		0.45%	0.68%	
30	KS	<i>PROKR2</i>	c.518T>G	p.Leu173Arg	Het	D	D	LOF	0.22%	0.35%	Reynaud 2012, JCEM
31	nCHH	<i>GNRHR</i>	c.350T>G	p.Leu117Arg	Het	D	D		private	private	Zhu 2015, JCEM
			c.266T>A	p.Leu89X	Het				private	private	
		<i>FGFR1</i>	c.1368G>T	p.Met456Ile	Het	D	T		0.04%	0.07%	Sykiotis 2010, PNAS
		<i>AXL</i>	c.1549G>A	p.Gly517Ser	Het	D	D		0.45%	0.68%	
32	nCHH	<i>KISS1</i>	c.349T>C	p.Phe117Leu	Het	T	T	LOF	0.02%	0.07%	Chan 2011, JCEM
33	nCHH	<i>KISS1</i>	c.154_156dupCCG	p.Pro52dup	Het				0.00%	0.00%	
34	KS	<i>KAL1</i>	c.256T>A	p.Cys86Ser	Hem	D	D		private	private	
		<i>FGFR1</i>	c.1961dupA	p.Tyr654X	Het				private	private	
35	KS	<i>CHD7</i>	c.8416C>G	p.Leu2806Val	Het	T	D		0.11%	0.09%	Bilan 2012, J Mol Diag
		<i>SOX10</i>	c.191A>T	p.Asp64Val	Het	D	T		0.00%	0.00%	
36	nCHH	<i>PROK2</i>	c.163delA	p.Ile55fs	Hom				0.01%	0.02%	Cole 2008, JCEM
37	KS	<i>FGF8</i>	c.77C>T	p.Pro26Leu	Het	T	T	LOF	0.04%	0.49%	Falardeau 2008, JCI
		<i>SOX10</i>	c.89C>A	p.Ser30X	Het				private	private	
38	KS	<i>PROKR2</i>	c.271C>T	p.Leu91Phe	Het	D	D		0.00%	0.00%	
39	nCHH	<i>HS6ST1</i>	c.652C>T	p.Pro218Ser	Het	T	D		0.16%	0.27%	
		<i>CHD7</i>	c.5051-4C>T		Het			+20.3%	0.54%	0.80%	
40	nCHH	<i>GNRH1</i>	c.141G>C	p.Glu47Asp	Het	T	D		0.14%	0.17%	
		<i>FGFR1</i>	c.622-1G>T		Het			-134.5%	private	private	

41	nCHH	<i>AXL</i>	c.1549G>A	p.Gly517Ser	Het	D	D		0.45%	0.68%	
42	KS	<i>FGFR1</i>	c.2233C>T	p.Pro745Ser	Het	D	D		private	private	Sato 2004, JCEM
43	nCHH	<i>FGFR1</i>	c.2464C>T	p.Arg822Cys	Het	D	D		0.02%	0.02%	
44	nCHH	<i>TACR3</i>	c.824G>A	p.Trp275X	Het				0.02%	0.04%	Gianetti 2010, JCEM
		<i>TAC3</i>	c.248A>G	p.His83Arg	Het	D	D		0.02%	0.02%	
45	KS	<i>FGFR1</i>	c.1093_1094dupAG	p.Pro366fs	Het				private	private	
		<i>CHD7</i>	c.8188G>A	p.Ala2730Thr	Het	T	D		0.00%	0.00%	
		<i>IL17RD</i>	c.2068T>A	p.Ser690Thr	Het	D	D		0.00%	0.01%	
46	nCHH	<i>GNRHR</i>	c.784C>T	p.Arg262Trp	Het	D	D	LOF	0.00%	0.00%	De Roux 1997, NEJM
			c.317A>G	p.Gln106Arg	Het	D	D	LOF	0.25%	0.40%	De Roux 1997, NEJM
47	KS	<i>FGFR1</i>	c.1038_1039insTT	p.Ile347fs	Het				private	private	
48	KS	<i>CHD7</i>	c.7282C>T	p.Arg2428X	Het				private	private	
49	KS	<i>FGFR1</i>	c.296A>G	p.Tyr99Cys	Het	D	D	LOF	private	private	Dodè 2003, Nat Genet; Raivio 2009, JCEM
50	KS	<i>PROKR2</i>	c.518T>G	p.Leu173Arg	Het	D	D	LOF	0.22%	0.35%	Reynaud 2012, JCEM
51	KS	<i>HS6ST1</i>	c.652C>T	p.Pro218Ser	Het	T	D		0.16%	0.27%	
		<i>CHD7</i>	c.2966G>A	p.Cys989Tyr	Het	D	D		private	private	
52	KS	<i>FGFR1</i>	c.790A>T	p.Asn264Tyr	Het	D	D		private	private	
53	KS	<i>CHD7</i>	c.5051-4C>T		Het			+20.3%	0.54%	0.80%	
54	nCHH	<i>FGFR1</i>	c.1306_1307dupTC	p.Met437fs	Het				private	private	
		<i>CHD7</i>	c.2613+5G>A		Het			-40.7%	0.00%	0.00%	

55	KS	<i>SOX10</i>	c.530G>A	p.Arg177Gln	Het	D	D		private	private	
56	KS	<i>CHD7</i>	c.3320C>T	p.Ala1107Val	Het	D	D		private	private	
57	nCHH	<i>CHD7</i>	c.3973T>C	p.Tyr1325His	Het	D	D		0.00%	0.01%	Bergman 2011, J Pediatr
58	nCHH	<i>FGFR1</i>	c.1552+1G>A		Het			-96.8%	private	private	
59	KS	<i>FGF8</i>	c.77C>T	p.Pro26Leu	Het	T	T	LOF	0.04%	0.49%	Falardeau 2008, JCI

Abbreviations as follows: KS, Kallmann syndrome; nCHH, normosmic congenital hypogonadotropic hypogonadism; Zyg, zygosity; Het, heterozygous; Hom, homozygous; Hem, hemizygous; D, deleterious; T, tolerated; PPH2, PolyPhen-2. PolyPhen-2 “possibly damaging” and “probably damaging” predictions were considered both as “deleterious”, while “benign” were defined as “tolerated” for consistency; LOF, loss-of-function, supporting deleteriousness of variant when both *in silico* algorithms predicted a tolerated effect.

Table S4. Number of screened individuals harboring mutated CHH genes.

# of genes mutated	CHH					KS					nCHH					CDGP				1000G	
	CHH	% CHH	p-value vs. CoLaus	p-value vs. 1000G	p-value vs. CDGP	KS	% KS	p-value vs. CoLaus	p-value vs. 1000G	p-value vs. CDGP	nCHH	% nCHH	p-value vs. CoLaus	P-value vs. 1000G	p-value vs. CDGP	CDGP	% CDGP	CoLaus	% CoLaus	1000G	% 1000G
0 genes	57	50.0%	5.5E-12	6.9E-14	7.7E-11	25	42.6%	6.8E-11	5.6E-13	5.4E-11	32	58.2%	1.3E-04	2.0E-06	3.0E-06	67	93.1%	333	82.2%	174	88.3%
1 gene	42	35.3%	5.8E-06	1.5E-07	5.8E-07	28	44.3%	4.6E-07	1.2E-08	3.7E-08	14	25.5%	ns	0.008	0.002	4	5.6%	64	15.8%	21	10.7%
≥2 genes	17	14.7%	6.4E-07	1.7E-06	0.002	8	13.1%	1.4E-05	2.0E-04	0.012	9	16.4%	3.0E-05	2.5E-05	0.002	1	1.4%	8	2.0%	2	1.0%

Number and frequency of cases and controls having no rare variants in CHH genes, one gene mutated or at least two genes mutated (oligogenicity). Differences between CHH, KS, and nCHH vs. CDGP probands and controls were analyzed via a two-sided Fisher's exact test.

Table S5. PTVs identified in CHH genes in patients and ExAC controls.

Gene	Inheritance	pLi	CHH (n=116)		ExAC NFE (n=33,370)	
			# of PTVs	Frequency	# of PTVs	Frequency
<i>IL17RD</i>	AD	0	0	0.0%	12	0.036%
<i>TACR3</i>	AR	0	1	0.9%	5	0.015%
<i>KISS1R</i>	AR	0	0	0.0%	4	0.012%
<i>PROKR2</i>	AR	0	0	0.0%	3	0.009%
<i>GNRHR</i>	AR	0.01	1	0.9%	1	0.003%
<i>TAC3</i>	AR	0.04	0	0.0%	1	0.003%
<i>NSMF</i>	AD	0.05	0	0.0%	4	0.012%
<i>WDR11</i>	AD	0.07	0	0.0%	10	0.030%
<i>PCSK1</i>	AR	0.09	0	0.0%	4	0.012%
<i>GNRH1</i>	AR	0.1	1	0.9%	3	0.009%
<i>PROK2</i>	AR	0.27	1	0.9%	2	0.006%
<i>FEZF1</i>	AR	0.3	0	0.0%	5	0.015%
<i>KISS1</i>	AR	0.52	0	0.0%	0	0.0%
<i>AXL</i>	AD	0.77	0	0.0%	6	0.018%
<i>FGF17</i>	AD	0.87	0	0.0%	1	0.003%
<i>HS6ST1</i>	AD	0.91	0	0.0%	0	0.0%
<i>SOX10</i>	AD	0.91	2	1.7%	0	0.000%
<i>FGF8</i>	AD	0.93	0	0.0%	0	0.000%
<i>ANOS1</i>	XLR	0.94	1	0.9%	2	0.006%
<i>SEMA3A</i>	AD	0.99	1	0.9%	4	0.012%
<i>FGFR1</i>	AD	0.99	9	7.8%	5	0.015%
<i>CHD7</i>	AD	1	3	2.6%	1	0.003%
<i>LEPR</i>	AR	1	0	0.0%	5	0.015%

Table S6. CNVs identified in CHH trios using array-CGH.

Pedigree	Phenotype	Genomic coordinates	CNV type	Genes in region	Inheritance
P313	KS	8:3699689-4005548	Deletion	<i>CSMD1</i>	From mother
P310	KS	1:246174091-246700678	Duplication	<i>SMYD3</i>	From father
		1:247083777-247416826	Duplication	<i>ZNF695, ZNF670, ZNF124</i>	
P327	KS	2:220185142-220322498	Duplication	<i>RESP18, DNPEP, DES, SPEG</i>	From mother
P312	nCHH	14:32110536-32257043	Deletion	<i>NUPBL</i>	From father

Table S7. Rare variants identified in family-based analyses.

Gene	Pedigree	Inheritance	Variant
<i>MGAT1</i>	P00163	DN	p.Arg129Trp
<i>TOP1MT</i>	P00163	DN	p.Gly419Arg
<i>BMX</i>	P00163	XLR	p.Glu422Ala
<i>DNAH1</i>	P00181	AR	p.Val1514Met/p.Asn2549Ser
<i>COL3A1</i>	P00181	DN	p.Gly459Arg
<i>SMC3</i>	P00181	DN	p.Cys549Tyr
<i>STS</i>	P00181	XLR	p.Pro517Ser
<i>CHRNA4</i>	P00184	AR	p.Pro554Leu/p.Glu92Gln
<i>PIEZO2</i>	P00184	AR	p.Arg1878His/p.Thr51Met
<i>CPXCR1</i>	P00184	XLR	p.Ala67Val
<i>FAM122B</i>	P00184	XLR	p.Ser207Pro
<i>RGAG1</i>	P00184	XLR	p.Ala839Thr
<i>ATRX</i>	P00267	XLR	p.Arg840Ile
<i>LCA10</i>	P00267	XLR	p.Gly148Arg
<i>PLOD2</i>	P00291	AR	p.Arg178His/p.Arg178His
<i>LRWD1</i>	P00310	AR	p.Val238Ile/p.Ala634Thr
<i>FAM47B</i>	P00310	XLR	p.Lys643Ile
<i>SLITRK2</i>	P00310	XLR	p.Ala760Ser
<i>LAMP2</i>	P00312	XLR	p.Gly221Arg
<i>PNMA5</i>	P00312	XLR	p.Arg96Cys
<i>RP2</i>	P00312	XLR	p.Asn19Lys
<i>CCDC64</i>	P00327	AR	p.Lys461Arg/p.Glu538Gln
<i>DUSP27</i>	P00327	AR	p.Arg493Trp/p.Phe1042Leu
<i>MACROD2</i>	P00327	AR	p.Gly314Arg/p.Gly314Arg
<i>NCOA1</i>	P00327	AR	p.Gly1276Arg/p.Ser1373Pro
<i>PKD1L2</i>	P00327	AR	p.Thr2145Met/p.Asp2063Glu
<i>ARSF</i>	P00327	XLR	p.Gly426Asp
<i>MAGEB1</i>	P00327	XLR	p.Tyr123Asn
<i>PHKA1</i>	P00327	XLR	p.Ala932Ser
<i>TFDP3</i>	P00327	XLR	p.Gln374Glu
<i>USP26</i>	P00327	XLR	p.Glu500Lys
<i>VMA21</i>	P00327	XLR	p.Arg39His
<i>BARHL1</i>	P00439	DN	p.Arg182Leu
<i>ERCC4</i>	P00439	DN	p.Arg740Cys
<i>RCAN1</i>	P00439	DN	p.Glu135Ala
<i>GUCY2F</i>	P00439	XLR	p.Gly434Arg
<i>WNK3</i>	P00439	XLR	p.Ala1305Val
<i>C11orf35</i>	P00454	AR	p.Glu463Val/p.Glu463Val
<i>DCHS2</i>	P00454	AR	p.Val1364Ala/p.Ala920fs
<i>DNHD1</i>	P00454	AR	p.Asn803Tyr/p.Ile4243Thr
<i>FAM193A</i>	P00454	AR	p.Ala12Val/p.Ala12Val
<i>GRIN3B</i>	P00454	AR	p.Val500Met/p.Ile706Val
<i>PKP3</i>	P00454	AR	p.Thr235Ser/p.Thr235Ser
<i>SLIT2</i>	P00454	AR	p.Asp1087His/p.Asp1087His
<i>NUP62CL</i>	P00454	XLR	p.Ser93Thr
<i>PLXNB3</i>	P00454	XLR	p.Val1275Ala
<i>TCEAL4</i>	P00454	XLR	p.Ala98Thr
<i>AIFM1</i>	P00489	XLR	p.Pro548Leu
<i>KLHL4</i>	P00489	XLR	p.Ala202Thr

<i>PGRMC1</i>	P00489	XLR	p.Gly174Arg
<i>TEX11</i>	P00489	XLR	p.Thr359Met
<i>TLR7</i>	P00489	XLR	p.Leu988Ser
<i>ARAP2</i>	P00827	AR	p.Glu753Gln/p.Asp743Tyr
<i>FAT4</i>	P00921	AR	p.Ile728Met/p.Glu4255Lys
<i>GUF1</i>	P00921	AR	p.Ala353Val/p.Ile478Ser
<i>ASAP3</i>	P00921	DN	p.Thr97Ala
<i>APEX2</i>	P00921	XLR	p.His269Tyr
<i>BCORL1</i>	P00921	XLR	p.Asp94Asn
<i>TIE1</i>	P00994	AR	p.Arg415His/p.Ser501Asn

AR: autosomal recessive; DN: de novo; XLR: X-linked recessive.

Table S8. RNA-seq results in candidate genes from family-based analysis.

Gene	Expression (voom)				Migrating GnRH neurons vs. nose		Post-migrating GnRH neurons vs. brain		Migrating vs. Post-migrating GnRH neurons		Nose vs. brain	
	Migrating GnRH neurons	Post-migratory GnRH neurons	Nose	Brain	logFC	adj. P value	logFC	adj. P value	logFC	adj. P value	logFC	adj. P value
<i>ASAP3</i>	0.00	0.22	0.15	0.06	-0.94	0.35	-1.01	0.57	0.49	0.48	0.97	0.66
<i>BARHL1</i>	0.00	0.00	0.00	0.00	-	-	-	-	-	-	-	-
<i>C11orf35</i>	0.00	0.00	0.70	0.00	0.64	0.66	-0.46	0.72	0.04	0.95	-2.17	0.22
<i>ERCC4</i>	1.53	1.62	3.91	0.88	0.57	0.92	1.78	0.58	-0.17	0.94	-0.62	0.90
<i>GUF1</i>	30.16	10.96	26.32	17.53	-0.47	0.87	0.46	0.93	-2.41	0.45	-1.15	0.57
<i>MGAT1</i>	6.51	3.23	5.00	2.03	0.18	0.96	-0.67	0.84	0.31	0.93	-0.73	0.74
<i>POLR3B</i>	33.08	9.23	7.44	9.90	-2.51	0.33	-0.24	0.96	-3.01	0.13	-0.22	0.95
<i>RCAN1</i>	2.84	1.07	0.73	0.01	-0.26	0.96	-1.02	0.54	-0.54	0.84	-1.19	0.61
<i>SLIT2</i>	7.03	41.27	31.40	25.44	2.96	0.13	-1.38	0.39	4.30	0.02	-0.63	0.68
<i>SMC3</i>	69.91	49.57	57.16	69.36	-0.70	0.45	0.82	0.40	-0.90	0.43	0.63	0.20
<i>TOP1MT</i>	0.51	5.38	0.24	1.49	-0.13	0.97	-2.00	0.53	2.28	0.27	0.60	0.83

Supplementary References

Felix, T.M., Hanshaw, B.C., Mueller, R., Bitoun, P., and Murray, J.C. (2006). CHD7 gene and non-syndromic cleft lip and palate. *Am J Med Genet A* 140, 2110-2114.

Miraoui, H., Dwyer, A.A., Sykiotis, G.P., Plummer, L., Chung, W., Feng, B., Beenken, A., Clarke, J., Pers, T.H., Dworzynski, P., et al. (2013). Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. *Am J Hum Genet* 92, 725-743.

Reynaud, R., Jayakody, S.A., Monnier, C., Saveanu, A., Bouligand, J., Guedj, A.M., Simonin, G., Lecomte, P., Barlier, A., Rondard, P., et al. (2012). PROKR2 variants in multiple hypopituitarism with pituitary stalk interruption. *The Journal of clinical endocrinology and metabolism* 97, E1068-1073.

Hanchate, N.K., Giacobini, P., Lhuillier, P., Parkash, J., Espy, C., Fouveaut, C., Leroy, C., Baron, S., Campagne, C., Vanacker, C., et al. (2012). SEMA3A, a gene involved in axonal pathfinding, is mutated in patients with Kallmann syndrome. *PLoS Genet* 8, e1002896.

Balasubramanian, R., Choi, J.H., Francescato, L., Willer, J., Horton, E.R., Asimacopoulos, E.P., Stankovic, K.M., Plummer, L., Buck, C.L., Quinton, R., et al. (2014). Functionally compromised CHD7 alleles in patients with isolated GnRH deficiency. *Proceedings of the National Academy of Sciences of the United States of America* 111, 17953-17958.

Pitteloud, N., Meysing, A., Quinton, R., Acierno, J.S., Jr., Dwyer, A.A., Plummer, L., Fliers, E., Boepple, P., Hayes, F., Seminara, S., et al. (2006). Mutations in fibroblast growth factor receptor 1 cause Kallmann syndrome with a wide spectrum of reproductive phenotypes. *Molecular and cellular endocrinology* 254-255, 60-69.

Guran, T., Tolhurst, G., Bereket, A., Rocha, N., Porter, K., Turan, S., Gribble, F.M., Kotan, L.D., Akcay, T., Atay, Z., et al. (2009). Hypogonadotropic hypogonadism due to a novel missense mutation in the first extracellular loop of the neurokinin B receptor. *The Journal of clinical endocrinology and metabolism* 94, 3633-3639.

Meysing, A.U., Kanasaki, H., Bedecarrats, G.Y., Acierno, J.S., Jr., Conn, P.M., Martin, K.A., Seminara, S.B., Hall, J.E., Crowley, W.F., Jr., and Kaiser, U.B. (2004). GNRHR mutations in a woman with idiopathic hypogonadotropic hypogonadism highlight the differential sensitivity of luteinizing hormone and follicle-stimulating hormone to gonadotropin-releasing hormone. *The Journal of clinical endocrinology and metabolism* 89, 3189-3198.

Gianetti, E., Tusset, C., Noel, S.D., Au, M.G., Dwyer, A.A., Hughes, V.A., Abreu, A.P., Carroll, J., Trarbach, E., Silveira, L.F., et al. (2010). TAC3/TACR3 mutations reveal preferential activation of gonadotropin-releasing hormone release by neurokinin B in neonatal life followed by reversal in adulthood. *The Journal of clinical endocrinology and metabolism* 95, 2857-2867.

Bailleul-Forestier, I., Gros, C., Zenaty, D., Bennaceur, S., Leger, J., and de Roux, N. (2010). Dental agenesis in Kallmann syndrome individuals with FGFR1 mutations. *Int J Paediatr Dent* 20, 305-312.

Zhu, J., Choa, R.E., Guo, M.H., Plummer, L., Buck, C., Palmert, M.R., Hirschhorn, J.N., Seminara, S.B., and Chan, Y.M. (2015). A shared genetic basis for self-limited delayed puberty and idiopathic hypogonadotropic hypogonadism. *The Journal of clinical endocrinology and metabolism* 100, E646-654.

Sykiotis, G.P., Plummer, L., Hughes, V.A., Au, M., Durrani, S., Nayak-Young, S., Dwyer, A.A., Quinton, R., Hall, J.E., Gusella, J.F., et al. (2010). Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. *Proceedings of the National Academy of Sciences of the United States of America* 107, 15140-15144.

Chan, Y.M., Butler, J.P., Pinnell, N.E., Pralong, F.P., Crowley, W.F., Jr., Ren, C., Chan, K.K., and Seminara, S.B. (2011). Kisspeptin resets the hypothalamic GnRH clock in men. *The Journal of clinical endocrinology and metabolism* 96, E908-915.

Bilan, F., Legendre, M., Charraud, V., Maniere, B., Couet, D., Gilbert-Dussardier, B., and Kitzis, A. (2012). Complete screening of 50 patients with CHARGE syndrome for anomalies in the CHD7 gene using a denaturing high-performance liquid chromatography-based protocol: new guidelines and a proposal for routine diagnosis. *J Mol Diagn* 14, 46-55.

Cole, L.W., Sidis, Y., Zhang, C., Quinton, R., Plummer, L., Pignatelli, D., Hughes, V.A., Dwyer, A.A., Raivio, T., Hayes, F.J., et al. (2008). Mutations in prokineticin 2 and prokineticin receptor 2 genes in human gonadotrophin-releasing hormone deficiency: molecular genetics and clinical spectrum. *The Journal of clinical endocrinology and metabolism* 93, 3551-3559.

Falardeau, J., Chung, W.C., Beenken, A., Raivio, T., Plummer, L., Sidis, Y., Jacobson-Dickman, E.E., Eliseenkova, A.V., Ma, J., Dwyer, A., et al. (2008). Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *The Journal of clinical investigation* 118, 2822-2831.

Sato, N., Katsumata, N., Kagami, M., Hasegawa, T., Hori, N., Kawakita, S., Minowada, S., Shimotsuka, A., Shishiba, Y., Yokozawa, M., et al. (2004). Clinical assessment and mutation analysis of Kallmann syndrome 1 (KAL1) and fibroblast growth factor receptor 1 (FGFR1, or KAL2) in five families and 18 sporadic patients. *The Journal of clinical endocrinology and metabolism* 89, 1079-1088.

de Roux, N., Young, J., Misrahi, M., Genet, R., Chanson, P., Schaison, G., and Milgrom, E. (1997). A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *The New England journal of medicine* 337, 1597-1602.

Dode, C., Levilliers, J., Dupont, J.M., De Paepe, A., Le Du, N., Soussi-Yanicostas, N., Coimbra, R.S., Delmaghani, S., Compain-Nouaille, S., Baverel, F., et al. (2003). Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat Genet* 33, 463-465.

Raivio, T., Sidis, Y., Plummer, L., Chen, H., Ma, J., Mukherjee, A., Jacobson-Dickman, E., Quinton, R., Van Vliet, G., Lavoie, H., et al. (2009). Impaired fibroblast growth factor receptor 1 signaling as a cause of normosmic idiopathic hypogonadotropic hypogonadism. *The Journal of clinical endocrinology and metabolism* 94, 4380-4390.

Bergman, J.E., Bocca, G., Hoefsloot, L.H., Meiners, L.C., and van Ravenswaaij-Arts, C.M. (2011). Anosmia predicts hypogonadotropic hypogonadism in CHARGE syndrome. *J Pediatr* 158, 474-479.